Platform for the production of cyclic and aromatic compounds in Saccharomyces cerevisiae

Kromphardt, Kresten Jon Korup

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Technical University of Denmark

Platform for the production of cyclic and aromatic compounds in *Saccharomyces cerevisiae*

PhD thesis by

Kresten Jon Korup Kromphardt

28^{th} of March 2019

DTU Bioengineering
Department of Biotechnology and Biomedicine

Platform for the production of cyclic and aromatic compounds in Saccharomyces cerevisiae

Kresten Jon Korup Kromphardt, 28^{th} March 2019

Address:
Technical University of Denmark
Department of Biotechnology and Biomedicine
Sølftofts Plads, Building 223
2800 Kgs. Lyngby
Denmark

Supervisors:
Professor Thomas Ostenfeld Larsen
Associate Professor Rasmus John Norman Frandsen

Assesment committee:
Professor Thomas James Simpson, University of Bristol
Senior Department Manager José Arnau, Novozymes
Associate Professor Ling Ding, Technical University of Denmark

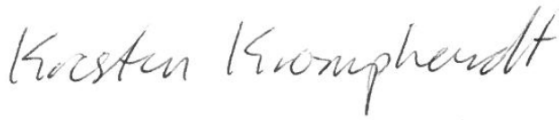
Date of defence:
27^{th} May - 2019

Preface

This thesis is part of a fulfillment of the requirements to obtain a PhD degree from Technical university of Denmark (DTU). The work presented here has been supervised by Professor Thomas Ostenfeld Larsen and Associate professor Rasmus John Norman Frandsen from March 2016 to March 2019. The work was initiated at DTU Department of Systems Biology and following some organisational changes within biosciences at DTU the name of the department was changed to DTU Department of Bioengineering and Biomedicine. The work was funded by Novo Nordisk foundation within the application round for Biotechnology-based Synthesis and production (grant number NNF15OC0016626).

During my PhD project I had the opportunity to visit the Yi Tang lab at University of California in Los Angeles. The stay was from August to November of 2017 and the focus of this stay was to learn techniques within enzyme production and purification, and *in vitro* assays within secondary metabolism.

Kgs. Lyngby, March 2019



Kresten Jon Korup Kromphardt

Contents

Acknowledgements

After three years of studying for my PhD degree at DTU Bioengineering there are many people I would like to give my warmest appreciation, as the past three years would not have been nearly as enjoyable or even possible without their help, encouragements and feedback.

First and foremost I would like to thank my supervisors Professor Thomas Ostenfeld Larsen and Associate professor Rasmus John Norman Frandsen for confidently leading me to where I am now. Their right (tough) questions have been pivotal to the development of me as a researcher and the PhD project would not have been possible without their guidance and encouragement throughout my three years.

I would like to thank the wonderful people at the Section for Synthetic biology for making me want to come to work everyday as it has always been a place of warm laughs, great nerdy comedy, exciting science and support when needed. I would like to thank Christine Alexandra Sværke Skovbjerg for fruitful discussions and collaboration throughout my PhD project. I would also like to thank Hilde Coumou who helped me in understanding yeast molecular biology. Finally, I would like to thank Peter Wolff for thoughtful discussion and I have the impression that both of us actually got wiser through the process, though others may disagree.

I have been a heavy user of the DTU Bioengineering Metabolomics Core during my PhD project, and many times I have been in the need of help. Therefore, I would also like to give gratitude to Christopher Phippen and Andreas Heidemann and recognize them for always being ready to give support when needed. Furthermore, I would like to than Christopher for giving me a thorough introduction to HPLC-MS quantification.

Erwin Schoof and Lene Holberg Blicher of the Proteomics Core were extremely helpful in giving me an introduction to protein extraction, sample preparation and proteomics data analysis and I would like to express my warmest gratitude for letting me be part of their everyday life for several weeks.

During my PhD I was very fortunate to have an external stay with Professor Yi Tang at UCLA for three months. During the stay Professor Tang and his rest of the wonderful people at his lab graciously shared their expertise within fungal secondary metabolism with me and I would like to express my warmest appreciation to Professor Tang and everyone at the Yi Tang lab at UCLA.

I would also like to thank my friends and family for their curiosity in my project

and support when times were tough.

Finally, Sigrid who has brought me love and appreciation, even when I found myself tumbling down the rabbit hole and got home hours later than I had anticipated. Even, then she has always brought a smile upon my face.

Summary

The switch from a petrol based economy to a biobased economy has brought researchers to investigate how biotechnology may be utilized to create cell factories producing already known compounds, while production of new-to-nature or designed compounds are also receiving attention. The understanding of natural product chemistry has long been improving, especially with the application of molecular biology to investigate the enzymatic machinery behind the biosynthesis of natural products. The understanding of polyketide synthases, which produces the natural products polyketides, has also brought about engineering attempts of these complex enzymes and it has to some degree been possible to design novel polyketide scaffolds. The goal of this PhD project was to investigate the possibility to create a programmable platform for creation of polyketides with a given chain length and folding pattern.

Initially, the possibility of combining type III PKSs with cyclases from type II PKS systems with *Saccharomyces cerevisiae* as production organism was investigated (Chapter 3 - Manuscript I). This approach was an attempt to divide the polyketide chain synthesis and the folding of the polyketide chain, into different enzymes making combinatorial experiments possible. It was found that a variety type III PKSs were functionally expressed in *S. cerevisiae* producing polyketide chains of C_6 to C_{16} depending on the expressed PKS. With the combination of the C_{16} producing octaketide synthase (OKS) and two different cyclases directing either the C7-C12 or C9-C14 folding patterns, it was shown that it was possible to direct the folding of the type III PKS polyketide product. This was a proof-of-concept step towards making a platform to design polyketides with a certain chain length and folding pattern.

Next, we wanted to investigate if it was possible to utilize the knowledge gained in the initial study to produce flavokermesic acid, which is a precursor of the industrial colorant carminic acid (Chapter 4 - Manuscript II). The pathway to produce flavokermesic acid consisting of a type III PKS and two cyclases had previously been expressed in *Nicotiana benthamiana* and *Aspergillus nidulans*. The aim of this study was to investigate if utilizing *S. cerevisiae* as a cell factory production host was feasible, as well as investigating if fusion of biosynthetic enzymes would limit shunt products of the pathway, and thereby improve the production of flavokermesic acid. It was found that *S. cerevisiae* was able to produce 52 mg flavokermesic acid per liter of medium if the biosynthetic enzymes were not fused,

while all versions of the fused biosynthetic enzymes were found to have reduced levels of flavokermesic acid, compared to the non-fused version. Two more tailoring steps are needed in order to synthesize carminic acid from flavokermesic acid, but the work presented here shows the potential of the platform to create a polyketide scaffold, which can be developed into an industrially important compound.

For the platform to be fully developed the incorporation of tailoring enzymes to alter the polyketide scaffolds are needed. To gain insight into a natural system, evolved to generate a high chemical diversity of polyketides, investigation of the *pgl1* gene cluster in the fungal genus *Fusarium* was conducted (Chapter 5 - Manuscript III). The investigation had two aims. The first being to investigate the diversity of the biosynthetic gene cluster throughout the genus of *Fusarium*, and the second being to propose a biosynthetic model for the chemical diversity thought to be derived form the *pgl1* biosynthetic cluster. The genes in the biosynthetic gene cluster were found to be highly conserved across the genus while the topology of the cluster was found to be less conserved, especially in the species complexes solani, decemcellulare and buxicola. It was proposed that these species within these species complexes could contain more biosynthetic activities in the *pgl1* cluster than other species in the *Fusarium* genus. This was an interesting hypothesis as the biosynthetic steps needed to create the chemodiversity found in the putative biosynthetic pathway, included many more enzymatic activities than what is included in currently identified biosynthetic cluster. This could either mean that more genes are implicated in the biosynthetic pathway, that some of the needed conversions happen spontaneously, or that the biosynthetic enzymes are promiscuous and therefore have low substrate specificity and may have several enzymatic specificities.

In summary, this PhD project has investigated the possibilities of creating a programmable platform for production of aromatic polyketides in *S. cerevisiae*, which allow control of chain length and folding pattern. Furthermore it has investigated what enzymatic machinery is employed in nature to derivatize a given polyketide scaffold. This work lays the foundation for further development of the platform to expand polyketide diversity with longer polyketide chains and cyclases to direct more of the folding of the polyketide chain, and furthermore, to attempt introduction of tailoring enzymes to fully develop the chemodiversity, which is possible to create with the platform.

Sammenfatning

Overgangen fra en fossilbaseret økonomi til en biobaseret økonomi har ledt forskere til at undersøge, hvordan bioteknologi kan udnyttes til at skabe cellefabrikker, der producerer allerede kendte forbindelser, mens produktionen af new-to-nature eller designede forbindelser også modtager opmærksomhed. Kendskabet til naturstofkemi har længe været i fremgang, især via anvendelse af molekylærbiologi til at undersøge det enzymatiske maskineri bag biosyntesen af naturstoffer. Forståelsen af polyketidsyntaser, som producerer naturstofferne polyketider, har også medført ingeniørmæssige tiltag for at manipulere disse komplekse enzymer, og det har i nogen grad været muligt at designe nye polyketid afledte stoffer. Målet med dette Ph.D.-projekt var at undersøge muligheden for at skabe en programmerbar platform til dannelse af polyketider med en given kædelængde og foldemønster.

Indledningsvis blev muligheden for at kombinere type III PKS'er med cyklaser fra type II PKS systemer, med *Saccharomyces cerevisiaese* som produktionsorganisme, undersøgt (Kapitel 3 - Manuskript I). Denne fremgangsmåde var et forsøg på at opdele polyketidkædesyntesen og foldningen af polyketidkæden i forskellige enzymer, hvilket ville muliggøre kombinatoriske eksperimenter. Det blev fundet, at en række type III PKS'er var funktionelt udtrykt i *S. cerevisiae*, og producerende polyketidkæder af C_6 til C_{16} afhængigt af den udtrykte PKS. Ved kombination af den C_{16} producerende oktaketid syntase (OKS) og to forskellige cyklaser, der styrerede enten C7-C12 eller C9-C14 foldemønstre, blev det vist, at det var muligt at styre foldningen af type III PKS polyketidproduktet. Dette var et proof-of-concept mod at skabe en platform til at designe polyketider med en given kædelængde og foldemønster.

Dernæst ønskede vi at undersøge, om det var muligt at udnytte den viden, der blev gundlagt i det første studie til at fremstille flavokermessyre, som er et forstadie til det industrielle farvestof karminsyre (Kapitel 4 - Manuskript II). Den syntetiske biosyntesevej til fremstilling af flavokermesyre bestående af en type III PKS og to cyclaser havde tidligere blevet udtrykt i *Nicotiana benthamiana* og *Aspergillus nidulans*. Formålet med dette studie var at undersøge, om udnyttelse af *S. cerevisiaeas* som cellefabrik til netop denne biosyntesevej var muligt, samt at undersøge, om fusion af de biosyntetiske enzymer ville begrænse side produkterne fra biosyntesevejen og dermed hæve produktionen af flavokermesyre. Det blev konstateret, at *S. cerevisiae* producerede 52 mg flavokermesyre per liter medie, hvis de biosyntetiske enzymer ikke var fusioneret, imens alle versioner af de

fusionerede biosyntetiske enzymer blev fundet til at reducere niveauet af flavokermesyre sammenlignet med den ikke fusionerede version. Der er behov for to yderligere biosyntese trin for at syntetisere karminsyre fra flavokermesyre, men det nuværende studie viste potentialet i platformen til at producere et polyketid, der kan udvikles til et industrielt vigtigt stof.

For at platformen kan blive fuldt udviklet, er det nødvendigt at inkorporere enzymer til at danne deriviater af polyketid skeletterne. For at få indsigt i et naturligt system, der har udviklet sig til at generere en høj kemisk diversitet af polyketider, blev en undersøgelse af *pgl1* gen clusteret i svampeslægten *Fusarium* gennemført (Kapitel 5 - Manuskript III). Undersøgelsen havde to formål. Det første var at undersøge mangfoldigheden af det biosyntetiske gen cluster i *Fusarium* slægten, og det andet formal var at foreslå en biosyntetisk model for den kemiske mangfoldighed, der menes at være afledt af det biosyntetiske cluster *pgl1*. Generne i den biosyntetiske cluster viste sig at være stærkt konserverede gennem hele slægten, mens clusterens topologi viste sig at være mindre konserveret, især i artskomplekserne *solani*, *decemcellulare* og *buxicola*. Derfor blev det foreslået, at arter indenfor disse artskomplekser kunne indeholde flere biosyntetiske gener i *pgl1* gen clusteret end andre arter i *Fusarium* slægten. Dette var en interessant hypotese, da de biosyntetiske trin, der var nødvendige for at skabe den kemodiversitet, der blev fundet i den foreslåede biosyntetiske vej, omfattede flere enzymatiske aktiviteter end hvad der kunne forklares med det nuværende identificerede biosyntetiske cluster. Dette kan enten betyde, at flere gener er impliceret i biosyntesen, at nogle af de nødvendige trin sker spontant eller at de biosyntetiske enzymer er promiskuøse og dermed har lav substratspecificitet og har flere enzymatiske specificiteter.

For at opsummere har dette Ph.D. projekt undersøgt mulighederne for at skabe en programmerbar platform til produktion af aromatiske polyketider i *S. cerevisiae*, som tillader kontrol af kædelængde og foldemønster. Endvidere har det undersøgt, hvilke enzymatiske mekanismer der anvendes i naturen til at derivatisere et givent polyketid skelet. Dette projekt ligger grundlaget for videreudvikling af platformen for at udvide polyketid diversiteten med længere polyketidkæder og flere cyclaser for at styre foldningen af polyketidkæden og endvidere at forsøge at indføre enzymer til deriviaterere polyketid skelettet og dermed at fuldt udvikle kemodiversiteten, det er mulig med platformen.

Publications

During the PhD project I have been conducting research, which have resulted in 3 manuscripts of which two are ready for submission and one is currently in preparation.

Manuscripts included in thesis

Steering polyketide production in yeast by combinatorial expression of enzymes from type II and III polyketide synthase systems

Kromphardt KJK, Skovbjerg CAS, Coumou HC, Larsen TO and Frandsen RJN

Manuscript currently in preparation

Production of flavokermesic acid in yeast by enzyme fusion

Kromphardt KJK, Skovbjerg CAS, Vestergaard AM, Rasmussen R, Hagerup T, Larsen TO and Frandsen RJN

Manuscript to be submitted to Scientific Reports

The *pgl1* biosynthetic gene cluster and metabiosynthetic pathway for fusarubins in *Fusarium*

Kromphardt KJK, Kim H-S, Proctor RH, Larsen TO and Frandsen RJN

Manuscript to be submitted to Fungal Genetics and Biology

Articles and conference abstracts not included in thesis

Metabolite production by species of *Stemphylium*

Olsen KJK, Rossman A & Andersen B

Fungal biology, **122**(2-3) pp 172-181, 2018, https://doi.org/10.1016/j.funbio.2017.12.012

Chemotaxonomy of the genus *Stemphylium*

Olsen KJK & Andersen B

Poster Presentation, *Joint Natural Product Conference*, June 2016

Microbial platform for production of aromatic compounds

Skovbjerg CAE, Olsen KJK, Larsen TO & Frandsen RJN

Poster Presentation, *Directing Biosynthesis V*, March 2017

***In vivo* type I iPKS engineering**

Olsen KJK, Larsen TO & Frandsen RJN

Poster Presentation, *Synthetic Biology – Engineering, Evolution and Design (SEED)*, June 2018

Abbreviations

Acetyl-CoA	Acetyl Coenzyme A
ACP	Acyl carrier protein
AT	Acyltransferase domain
DEBS	6-deoxyerythrnolide synthase
DH	Dehydratase domain
ER	Enoylreductase domain
FAS	fatty acid synthase
HR-PKS	Highly reducing PKS
HRMS	High resolution mass spectroscopy
iPKS	Iterative PKS
KR	Ketoreductase domain
KS	Ketosynthase domain
Malonyl-CoA	Malonyl Coenzyme A
mPKS	Modular PKS
NR-PKS	Non-reducing PKS
NRPS	Non-ribosomal peptide syntethase
PKS	Polyketidesynthase
PR-PKS	Partly reducing PKS
TE/CLC	Thioesterase or cyclisation domain
UHPLC	Ultra high performance chromatography
UV/Vis	Ultraviolet/Visual light

1 Project motivation and aim

1.1 Project motivation

As the issues of the petrol based economy has arisen and the options of switching to a biobased economy begun to be investigated. To reach this goal many new biological solutions have to be developed as the range of products and commodities derived from the petrochemical industry is large. These range from bulk chemicals such as fuels, over commodity chemicals such as plastics to fine chemicals such as pharmaceuticals [1, 2]. The development of cell factories to produce such compounds based only on biological materials are underway from many sides and focus is given on many different parts of the process such as diversification of possible growth substrate for production [3, 4], increasing output of a selected metabolic pathway by metabolic engineering [5] as well as development of the product range to replace more of the petroleum derived compounds [6, 7]. An important compound group within the fine chemicals are polyketides of which many are used as medicine [8] or as additives to food and cosmetics [9]. The realisation of the biobased economy requires the development of new compounds within all classes of chemicals and engineering efforts has created toolboxes to design specific polyketides [10, 11, 12].

1.2 Project aim and thesis outline

The overall goal of the work presented in this thesis was to develop a platform to produce aromatic polyketides in a programmable fashion in the industrial workhorse *Saccharomyces cerevisiae*. The choice of *S. cerevisiae* as the organism for the platform was that it had been used for the construction of many different biosynthetic pathways [13, 14] and amplification of a pathway of choice to improve production titers [15, 16]. One solution for creating a programmable biosynthetic pathway is to divide pathway product formation into distinct steps (modules) and then attempt to control them individually. In the case of producing aromatic polyketides the biosynthetic problem can be divided into the formation of the backbone polyketide chain, and secondly the folding of this chain to yield

the desired cyclic structure. A key concept in synthetic biology is modularity in design, and control over the two described steps would allow for production of a large array of polyketide backbones.

Specific aims for the given project were as follows:

1. Devise and provide proof-of-concept for a programmable platform that allows for the production of cyclic and aromatic polyketides in *S. cerevisiae*
2. Evaluate the potential for the production of a relevant industrial chemical utilizing the platform
3. Investigate how further derivation of polyketides are processed in nature

These aims have been addressed throughout the project, which have resulted in three manuscripts. The content of the thesis is outlined here:

Chapter 2 - Theoretical background of polyketide synthesis to understand the production of polyketides in nature, while also giving insight into past strategies to engineer various type of polyketide producing enzymatic machineries.

Chapter 3 - Describes the proof-of-concept for a programmable platform in *S. cerevisiae* to create cyclic and aromatic polyketides by combination of type III PKSs and cyclases from type II PKS systems.

Chapter 4 - Describes the establishment an artificial biosynthetic pathway in *S. cerevisiae* to produce flavokermesic acid, which is a precursor of the widely used industrial colorant carminic acid.

Chapter 5 - Investigation of a shared biosynthetic pathway the production of fusarubins and related compounds in the fungal genus *Fusarium*, and how the metabolites isolated from individual species can be connected through few biochemical conversions to create a more elaborate metabiosynthetic pathway. Furthermore, the biosynthetic gene cluster producing the metabolites and its distribution throughout the *Fusarium* genus is investigated.

Chapter 6 - Gives a general discussion of the thesis and draws the overall conclusions as well as giving some perspectives of the work presented in the thesis.

The references given inside each chapter is referring to the reference list, which is in the end of that particular chapter.

References

[1] Erickson, B and Winters, P. "Perspective on opportunities in industrial biotechnology in renewable chemicals". In: *Biotechnol J* 7.2 (2012), pp. 176–185.

[2] Sheldon, RA. "Green and sustainable manufacture of chemicals from biomass: state of the art". In: *Green Chem* 16.3 (2014), pp. 950–963.

[3] Chatzifragkou, A, Makri, A, Belka, A, Bellou, S, Mavrou, M, Mastoridou, M, Mystrioti, P, Onjaro, G, Aggelis, G, and Papanikolaou, S. "Biotechnological conversions of biodiesel derived waste glycerol by yeast and fungal species". In: *Energy* 36.2 (2011), pp. 1097–1108.

[4] Singla, A, Paroda, S, Dhamija, SS, Goyal, S, Shekhawat, K, Amachi, S, and Inubushi, K. "Bioethanol production from xylose: problems and possibilities". In: *J Biofuels* 3.1 (2012), pp. 39–49.

[5] Martel, CM, Warrilow, AGS, Jackson, CJ, Mullins, JGL, Togawa, RC, Parker, JE, Morris, MS, Donnison, IS, Kelly, DE, and Kelly, SL. "Expression, purification and use of the soluble domain of *Lactobacillus paracasei* β-fructosidase to optimise production of bioethanol from grass fructans". In: *Bioresour Technol* 101.12 (2010), pp. 4395–4402.

[6] Bond, CM and Tang, Y. "Engineering *Saccharomyces cerevisiae* for production of simvastatin". In: *Met Eng* 51 (2019), pp. 1–8.

[7] Kildegaard, KR, Wang, Z, Chen, Y, Nielsen, J, and Borodina, I. "Production of 3-hydroxypropionic acid from glucose and xylose by metabolically engineered *Saccharomyces cerevisiae*". In: *Met Eng Commun* 2 (2015), pp. 132–136.

[8] Gomes, ES, Schuch, V, and Lemos, EGdM. "Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics". In: *Braz J Microbiol* 44.4 (2013), pp. 1007–1034.

[9] Baranyovits, FLC. "Cochineal carmine: an ancient dye with a modern role". In: *Endeavour* 2.2 (1978), pp. 85–92.

[10] Newman, AG, Vagstad, AL, Storm, PA, and Townsend, CA. "Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming". In: *J Am Chem Soc* 136.20 (2014), pp. 7348–7362.

[11] McDaniel, R, Ebert-Khosla, S, Hopwood, DA, and Khosla, C. "Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits". In: *Nature* 375.6532 (1995), pp. 549–554.

[12] Frandsen, RJ, Khorsand-Jamal, P, Kongstad, KT, Nafisi, M, Kannangara, RM, Staerk, D, Okkels, FT, Binderup, K, Madsen, B, Møller, BL, Thrane, U, and Mortensen, UH. "Heterologous production of the widely used natural food colorant carminic acid in *Aspergillus nidulans*". In: *Sci Rep* 8.12853 (2018), pp. 1–10.

[13] Bond, C, Tang, Y, and Li, L. "*Saccharomyces cerevisiae* as a tool for mining, studying and engineering fungal polyketide synthases". In: *Fungal Genet Biol* 89 (2016), pp. 52–61.

[14] Albertsen, L, Chen, Y, Bach, LS, Rattleff, S, Maury, J, Brix, S, Nielsen, J, and Mortensen, UH. "Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes". In: *Appl Environ Microbiol* 77.3 (2011), pp. 1033–1040.

[15] Strucko, T, Buron, LD, Jarczynska, ZD, Nødvig, CS, Mølgaard, L, Halkier, BA, and Mortensen, UH. "CASCADE, a platform for controlled gene amplification for high, tunable and selection-free gene expression in yeast". In: *Sci Rep* 7.41431 (2017), pp. 1–12.

[16] Chen, Y, Daviet, L, Schalk, M, Siewers, V, and Nielsen, J. "Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism". In: *Met Eng* 15 (2013), pp. 48–54.

2 Background

The field of natural product chemistry is broad and has been through rapid development over the past decades as our understanding of the biosynthetic origin of natural products. This has been aided by a mechanistic understanding of the underlying genetic and enzymatic machinery, which is responsible for the production these complex biomolecules. Especially, the application of molecular genetic tools has sped up the elucidation of secondary metabolic pathways. The application of molecular techniques like gene knockout in the native host [1], expression of pathway encoding genes in heterologous hosts [2] and enzyme purification followed by *in vitro* enzymatic assays [3], has been the most powerful tools in getting a grip on natural product biosynthesis. Furthermore, chemical analysis methods have been in a rapid development. The application of high performance chromatography methods for better separation of complex natural samples, as well as development of more sensitive and precise detection methods such as high resolution mass spectroscopy (HRMS) detection, has been pivotal to the detection and elucidation of novel compounds [4].

Within secondary metabolism the four most important compound classes are terpenes, non-ribosomal peptides, shikimic acid derived compounds and polyketides. The vast majority of known secondary metabolites to date originate from one or more of these compound classes [5]. The first class, terpenes, are formed using C_5 isoprene units as building blocks, which are joined in the mevalonate pathway to form branched chains of highly reduced carbon chains. The name terpenes refers to its discovery in turpentine oil from pine trees. The chemical diversity of terpenes is created by the different terpene cyclase enzymes, which fold the isoprene derived chains into complex structures. The second compound class, non-ribosomal peptides, consist of small peptides, which are, as the name suggest not produced by ribosomes, but rather via an mRNA independant manner. Formation of the non-ribosomal peptide, are catalyzed by non-ribosomal peptide syntethase (NRPS), which join amino acids. The substrate for NRPSs are proteinogenic or non-proteinogenic amino acids, which are joined by peptide bonds as with regular protein synthesis in ribosomes, but cyclisation and other internal modifications make them unique in comparison with ribosomal peptides

and proteins. The third class originate from the important primary metabolite shikimic acid. The biosynthesis of aromatic amino acids depend on this metabolite and shikimic acid is also found in biosynthesis of secondary metabolites, as shikimic acid derived compounds from central metabolism can be used in secondary metabolism [5]. The fourth and final compound class, polyketides, have much structural similarity with fatty acids. The enzymatic machinery producing polyketides are called polyketide synthases and they utilize building blocks in the form of C_2 acetate units or derivatives thereof called ketides. The repeated joining of these building blocks yields a polyketide chain and is the foundation for polyketides. Fatty acids are mostly saturated or with very few double bonds, while polyketides can have many different levels of saturation [6, 7]. Even higher levels of diversity is obtained when two or more of the compound classes are combined to form hybrid natural products [8, 9].

These four compound classes are the foundation for production of secondary metabolites, but the diversity of secondary metabolism is only partly due to these four enzyme classes. After production of the basic chemical structures tailoring enzymes can carry out a plethora of extremely specific chemical reactions leaving organic synthesis chemists in envy. These can be reductions, oxidations, methylations or rearrangements to name a few [10].

The compound class of interest in this project has been polyketides and the engineering of the enzyme responsible for formation of these. Therefore, the following chapters will focus on introducing the mechanism of polyketides synthases in nature and thereafter give a review of previous attempts of engineering of polyketide synthases.

2.1 Production of polyketides in nature

Polyketides are an extremely diverse group of natural products. The diversity is not only in the range of known polyketide derived metabolites but also in the enzymes producing polyketides, namely the polyketide synthases (PKSs). There are three types of PKSs, I, II and III. In brief, the structure of type I PKSs consist of a single protein with a number of domains. Type II PKSs are composed of several individual enzymes, each with a unique function, which interact to produce the polyketide chain. Type III PKSs are homodimers consisting of only one domain [11, 12, 13].

Polyketide synthesis proceeds via the claisen condensation reaction that forms

C-C bonds between chain and incoming extender units. The formation of a new polyketide is initiated by the loading of a starter unit. The starter unit can be extremely diverse and examples include amino acids, fatty acids and many other, but the most prominent is acetyl coenzyme A (acetyl-CoA). Thereafter, claisen condensations, which are driven by elimination of CO_2 from the incoming extender substrate malonyl coenzyme A (malonyl-CoA) or derivatives thereof, elongate the polyketide chain. The elongation is continued until an enzyme specific chain length is reached, and the final product is released. The release can result in ring formation of the polyketide chain or simply by release of the polyketide chain from the PKS [7, 14].

The chemical diversity of polyketides are primarily determined by the biochemical activity harnessed by the domains in the PKS. The ketosynthase domain (KS) are common for all PKSs and is the only recognized domain in type III PKSs. In type I and II PKSs the acyl-carrier protein (ACP) is also an essential domain as it in tandem with the KS domain to accept elongation units from the acyltransferase domain (AT) and carries the growing polyketide chain. Also, the AT domain responsible for selection of the starter units in polyketide synthesis. Another important feature of the ACP is the ability to allow the growing polyketide chain to be modified by other domains in the PKS, which along with other hallmarks of the different PKS systems will be described in detail in the following sections. The general mechanism for polyketide synthesis and an overview of the different PKS systems is depicted in figure 2.1.

2.1.1 Type I modular PKSs

Type I polyketides are split into two groups, iterative and modular PKSs. The modular PKSs (mPKS) are almost exclusively found in bacteria and are the most intensively studied type of PKS. The reason for this interest is the important hallmark of these enzymes that the enzyme domain architecture and the resultant compound, which the enzyme produce, are correlated. This makes prediction of the final compound based on genetic information possible because each module of domains (KS, AT and ACP) in the mPKS is only used once, and therefore, the length and to a high degree other modifications (e.g. reductions or methylations) of the resultant polyketide chain can be predicted from the amino acid sequence of the mPKS alone. This principle is named colinearity [11, 13].

In between modules of the mPKS are a stretch of amino acids which have been

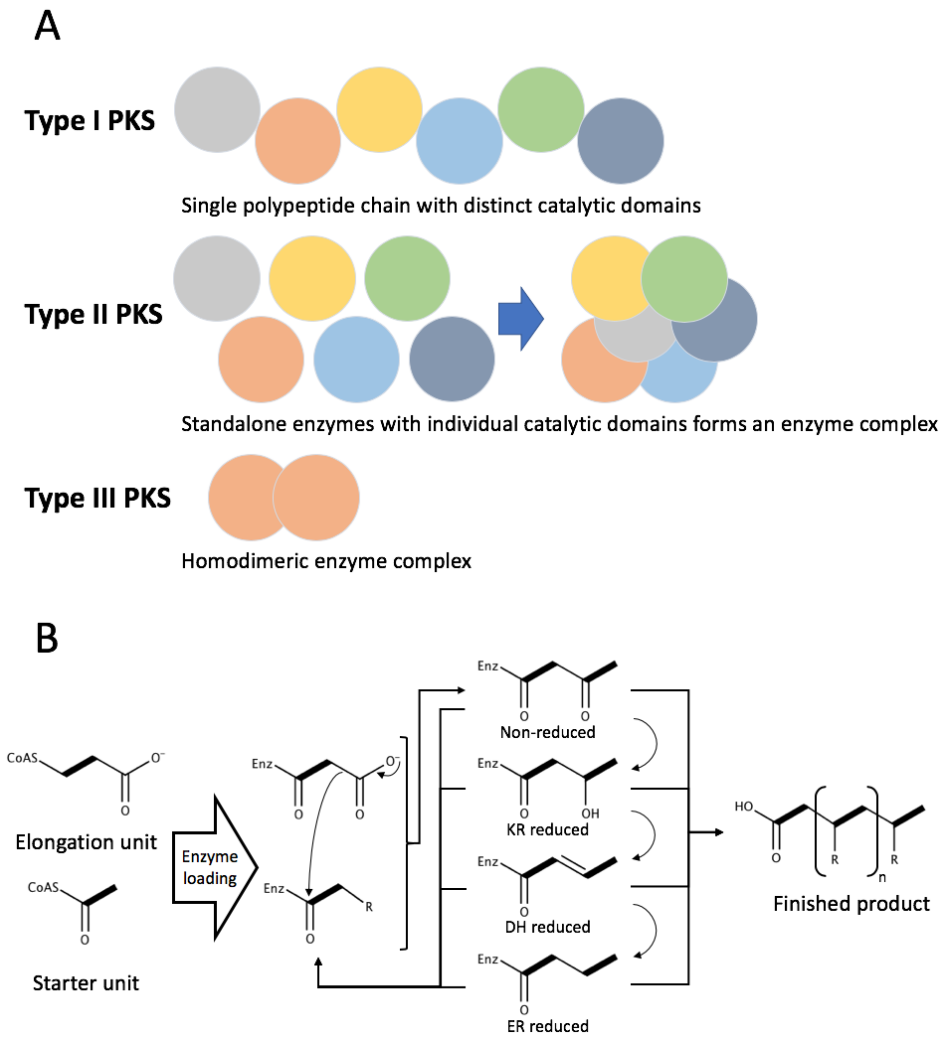


Figure 2.1 Overview of overall structure and mechanism of PKSs. (A) The structure of the different types of PKSs and hallmarks that make them distinguishable. (B) Generalized mechanism for polyketide synthesis. Synthesis is initiated by loading of the Coenzyme A activated starter unit onto the PKS, followed by loading of the elongation module. Elongation of the polyketide chain happens via claisen condensation. When the elongation has been done reduction domains may reduce the backbone if present in the PKS. After reduction, if needed another round of elongation is done. The polyketide synthesis is terminated by unloading of the polyketide chain which can also entail one or multiple ring formations of the polyketide chain.

named intralinkers and these allows the different modules of the mPKS to interact with each other and transport the growing polyketide chain through the mPKS while keeping the individual modules in close vicinity. An exception to the general rule of type I PKSs being a single polypeptide chain is found in mPKSs as these are often not expressed in a single polypeptide chain, but often as one or several of the modules in one chain. Then these chains of modules interact via what is named interlinker domains. These are domains which serve the same function as intralinkers, ie. to keep modules in close vicinity of one another, but does so through protein-protein interactions instead of the covalent bonds of the amino acid chain [15].

Reductions of the polyketide backbone while it is being synthesised are only known in type I PKSs. The reducing domains can alter the formed β-keto carbon chain into a partly or fully reduced chain. The initial reduction is carried out by the keto reductase (KR) domain and reduces the β-keto group to an alcohol. A second reduction can be carried out by the dehydratase (DH) domain and reduces the alcohol to a double bond. The third and final possible reduction is carried out by the enoyl reductase domain (ER), which eliminates the double bond and produces a fully reduced carbon chain as is known in many fatty acids where these three reducing domains are also present [5]. In type I mPKSs the level of reduction is specific for each module in the PKS. It is important to note that the reduction is carried out on the ketide unit fused to the chain by the upstream module to where the reducing domains are present. Therefore, when predicting the product from an mPKS, the modifications to the polyketide chain are dictated by the reducing domains in the downstream module. One exception from the colinearity principle in mPKSs are that sometimes domains are still present in a mPKS enzyme, e.g. reducing domains, but is inactive. This can be due to gradual mutations yielding some domains inactive, but evolution has not had enough time to completely delete the reducing domain or that the inactive domain is essential for the overall protein structure. When the final module has done the elongation of the polyketide chain a thioesterase domain (TE) unloads the chain which often also includes cyclisation of the chain to form macrolides [16].

The common product from type I mPKS are often polyketide backbones with very complex reduction patterns and just as often they utilize alternative elongation units to the otherwise often used malonyl-CoA. The most intensively studied mPKS is the 6-deoxyerythronolide synthase (DEBS), which produces a precursor

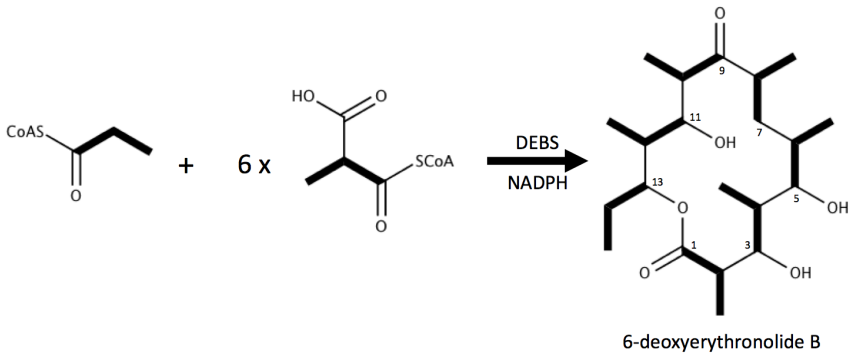


Figure 2.2 The biosynthesis of 6-deoxyerythronolide by DEBS.

to erythromycin and serves as a good example of this. First of all the starter unit is not an acetyl, but instead a propionyl is used as the initial unit of the final product 6-deoxyerythronolide. Subsequently six units of methylmalonyl-CoA are added to the initial propionyl starter unit. Furthermore, the reduction of the backbone is either to the fully reduced carbon backbone or to an alcohol. When the product of DEBS is unloaded an ester is formed between the terminal carboxylic acid of the polyketide chain and an alcohol, which was created by the second module via reduction of the keto group on the initial propionyl. This yields the macrolide 6-deoxyerythronolide, which is the first precursor of erythromycin (Figure 2.2) [15, 17].

2.1.2 Type I iterative PKSs

The type I iterative PKSs (iPKSs) are single a protein chain as is the case for mPKSs, however they only contain one copy of each domain, that are used repeatedly during synthesis to produce the polyketide. The type I iPKSs resemble the mechanism of type I FAS found in animals. Type I iPKSs are found almost exclusively in fungi, but examples in bacteria have been found [7, 11].

Type I iPKSs are split into three distinct groups based on their level of reduction, which are non-reducing PKSs (NR-PKSs), partly-reducing PKSs (PR-PKSs) and highly-reducing PKSs (HR-PKSs). As the names suggest the reducing ability of the PKs and thus the resultant product from the PKS dictates the naming of the PKS. The NR-PKSs produce backbones with only β-keto groups which can then be folded in different ways to generate complex aromatic structures. The folding carried out in NR-PKSs is directed by specific domains. The first ring

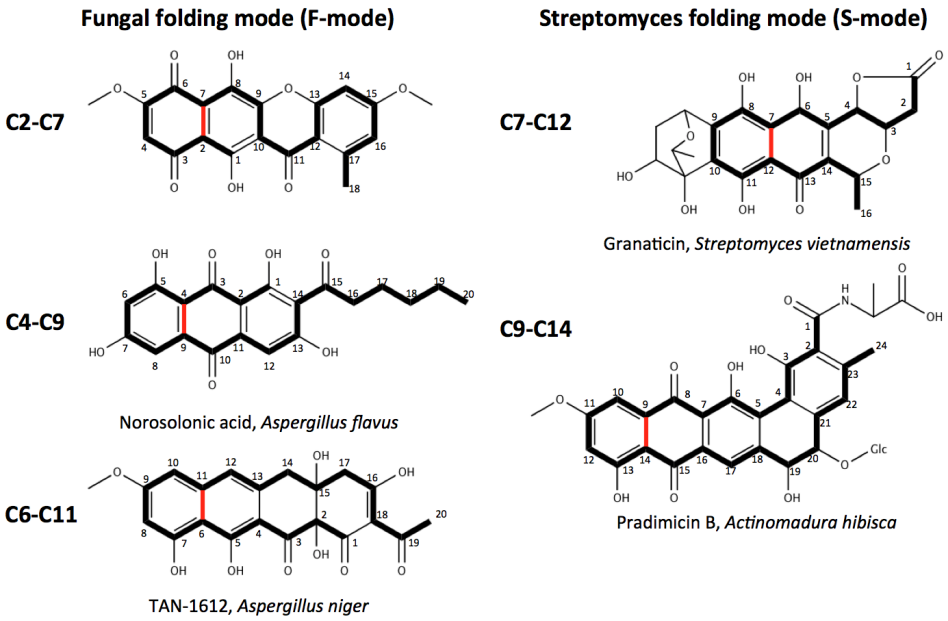


Figure 2.3 Examples of the F- and S-mode first ring folding in natural products, with the original polyketide backbone highlighted in black and the first ring formation highligted in red in each compound. Carbon number 1 is part of the last ketide unit being fused to the polyketide chain. F-mode examples: bikaverin [18], norosolonic acid [19], TAN-1612 [20]. S-mode examples: granaticin [21] and pradimicin (Glc = sugar moiety) [22].

folding is directed by the product template domain (PT). The folding patterns are determined by regarding the two carbons joined in the first ring closure and determining the distance of these to the terminal carbon added to the polyketide backbone, which usually is a carboxylic acid. Three possible folding patterns are most often found to be catalyzed by the PT domains being the C2-C7, C4-C9 and C6-C11 first ring folding patterns. It is important to note that this type of folding with the lower number being even has been termed Fungal-mode (F-mode) folding [23, 24] ,which is in contrast to *Streptomyces*-mode (S-mode) [25] folding (Figure 2.3), which will be explained in the following section of type II PKSs. The PT domain directs the first and often also the second ring formation while further ring formations may be directed by the TE domain, which in this case also acts as a cyclisation domain upon release. Up to four rings can be formed by the PKS itself as is seen in the biosynthesis of bikaverin and TAN-1612 (Figure 2.3) [7, 20, 26]. Our understanding of the NR-iPKSs domains has primarily been through the analysis of PksA, which is involved in the biosynthesis of aflatoxin in *Aspergillus flavus*.

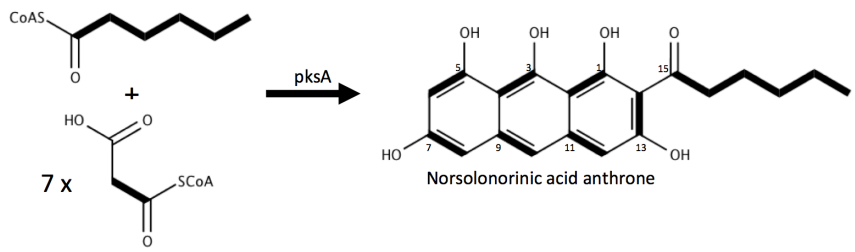


Figure 2.4 The biosynthesis of norsolorinic acid anthrone by PksA.

In short, the PKS-part of the biosynthesis of aflatoxin is initiated by loading of the starter unit hexanoyl-CoA, which is produced by the FAS enzymes HexA and HexB. This is followed by seven rounds of malonyl-CoA elongation of the chain, followed by PT catalyzed C4-C9 and C2-C11 ring formation and finally C1-C14 ring formation and unloading the norsolorinic acid anthrone product. The biosynthesis of norsolorinic acid is summarized in figure 2.4 [19, 27]. An alternative to the TE domain release is the reductive domain (R) release in which instead of a terminal carboxylic acid group the terminal carbon is an aldehyde [28].

The PR- and HR-iPKSs have some or all of the reducing domains described in the previous section on mPKSs present in the enzyme. The main difference between the two reducing PKSs are, as the name suggests, the level of reduction exerted on the polyketide backbone. PR-PKSs does not have all three reducing domains in their structure and therefore they produce polyketide chains which have double bonds or alcohol groups along with the ketone groups. 6-methyl salicylic acid synthase (MSAS) from *Penicillium patulum* is one of the most intensively studied PR-PKSs. The biosynthesis of 6-methyl salicylic acid (6-MSA) in *P. patulum* is initiated by loading of an acetyl-CoA starter unit followed by extensions with three malonyl-CoA units of which only the C5 carbonyl is reduced to an alcohol. The following C2-C7 ring closure catalyzed by the PT domain enables unloading of the 6-MSA. The biosyntheis of 6-MSA is shown in figure 2.5 [7].

The hallmark of HR-PKSs are the presence of all reducing domains in the PKS. This yields a polyketide backbone which cannot be folded by claisen condensation as in NR- and PR-PKSs, but cyclisation of these compounds may instead carried out via Diels-Alder cyclisation, as is seen in the biosynthesis of lovastatin [29]. The PKS LovB *Aspergillus terreus* producing the the precursor of lovastatin, dihydromonacholin L (Figure 2.6), is one of the most intensively studied HR-PKSs. In short, the biosynthesis of dihydromonacholin L is intiated by loading acetyl

as starter unit which is then elongated by eight malonyl-CoA units to yield a reduced nonaketide. The complex reduction machinery of LovB reduces several of the integrated ketide untis to double bonds and alcohol groups. Several of the ketides are fully reduced to single bonds by the trans acting ER LovC. Furthermore, the C-methylation domain of the PKS selectively methylates the C12 position. The characteristic double ring of lovastatin is formed via Diels-Alder ring formation and the product dihydromonacolin L, which is a precursor to lovastatin, is released from the PKS [29, 30].

As in mPKSs the level of reduction is determined by the enzymatic structure, but prediction of reduction is for now impossible for type I iPKSs. Predictions of the product from type I iPKSs based on sequence, can only be proposed via homology with other known iPKSs and the presence and absence of catalytic domains. Currently, the best predictive power of the product from an unknown iPKS is the folding pattern, which is catalyzed by the PT domain in NR-iPKSs [23]

Some polyketide products have vastly different oxidation levels in different parts of the backbone. Curvularin is an example of one such compound (Figure 2.7). In curvularin biosynthesis the first part of the synthesis is performed by a HR-PKS named CURS1 and the product is a highly reduced tetraketide. This tetraketide is then used as a starter unit for a NR-PKS CURS2 which carries out an additional three elongations of the polyketide and forms the final product curvularin. This polyketide is also one of the only known examples of a C3-C8 first ring formation by a PT domain. This exception to the otherwise well described F-mode folding was investigated by Xu *et al.* (2013). They found that two point mutations switched the PT of CURS2 from the uncommon C3-C8 folding pattern to the regular C2-C7 pattern [31]. Another example of variable oxidation levels are found in polyenes like aurovertins and citreoviridins (Figure 2.7). The biosynthesis

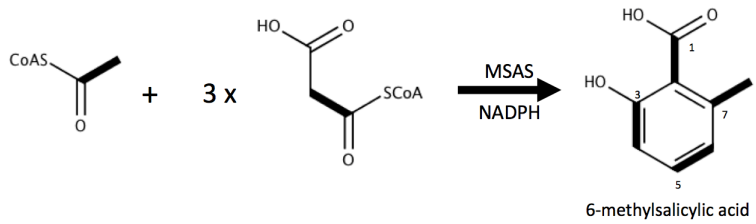


Figure 2.5 The biosynthesis of 6-methyl salicylic acid by MSAS.

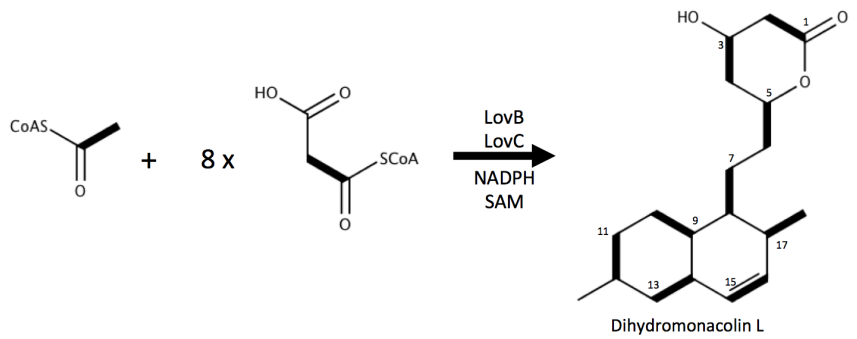


Figure 2.6 The biosynthesis of dihydromonacholin L by LovB and LovC.

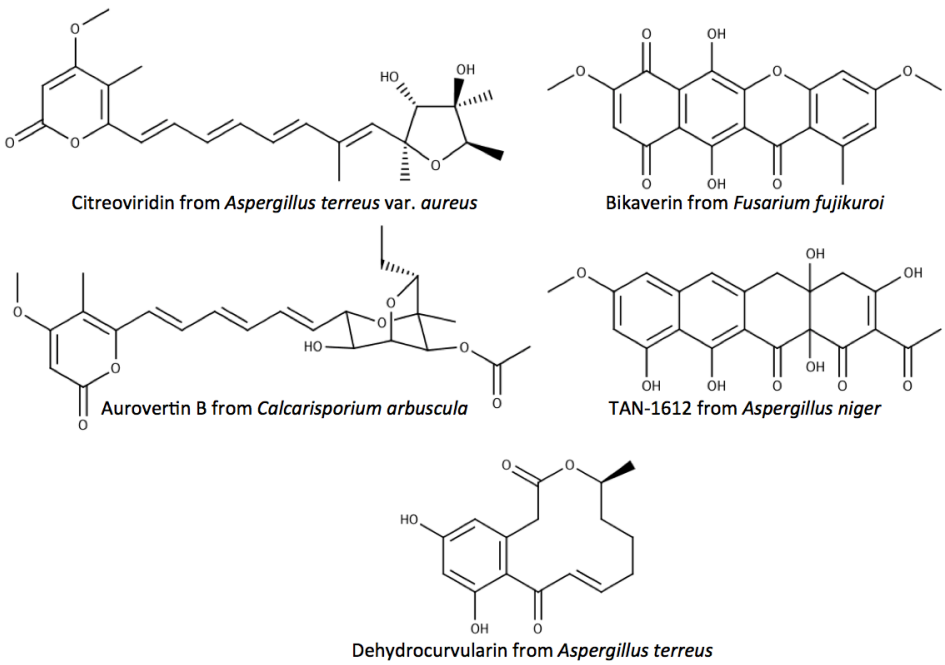


Figure 2.7 Compounds derived from type I iPKS systems.

of aurovertins and citreoviridins are carried out by one PKS (AurA and CtvA respectively), which is in contrast to biosynthesis of curvularin. The middle part of aurovertins and citreoviridins are polyenes and the terminal three elongations of the polyketide chain is non-reduced β-ketones. This suggest that the access to the DH reducing domain is dependent on the length of the polyketide chain [32, 33].

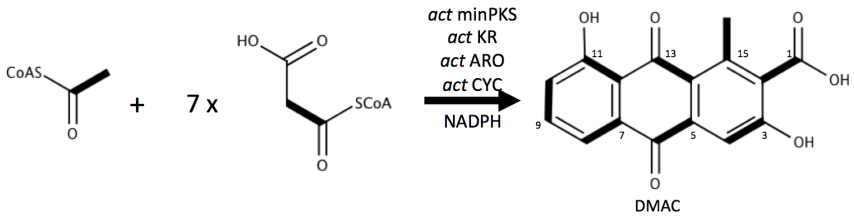


Figure 2.8 The biosynthesis of 3,8-dihydroxy-1-methylanthraquinone-2-carboxylic acid by the actionorhodin PKS.

2.1.3 Type II PKSs

As described earlier the strong contrast between type I and type II PKSs is that the catalytic domains of type II PKSs are located on individual enzymes instead of being part of the same protein. Another interesting difference is that type II PKSs are exclusively found in bacteria with actinobacteria having the vast majority of characterized type II PKSs [14, 34]. The malonyl-acetylation domain (MAT) loads malonyl onto two KS domains named KSα and KSβ[1], which forms a heterodimer. The KS domains together with the ACP have long been named the minimal PKS as these three are able to form polyketide products [35, 36]. With this being said these minimal PKS products are often shunt products as more enzymes are needed for the formation of the native product. These enzymes act upon the poly-β-keto chain to fold it in a certain shape and are named cyclases or aromatases. A subclass of these only act upon partly reduced ketide chains that have a C-9 alcohol group instead of the ketone [11, 35]. The most widely studied and best understood type II PKS system is from *Streptomyces coelicolor* and produces actinorhodin. The function of the actinorhodin PKS to form the actinorhodin precursor 3,8-dihydroxy-1-methylanthraquinone-2-carboxylic acid (DMAC) is based on elongation of the acetyl-CoA starter unit with seven malonyl-CoA units to create the C_{16} backbone. The backbone is reduced at the C9 position and C-C bonds are formed at C7-C12 and C5-C14 by trans acting KR and cyclase enzymes. Upon release the last C2-C15 bond is formed spontaneously to yield DMAC (Figure 2.8) [37, 38].

As mentioned earlier in NR-iPKSs there is a certain folding pattern observed for type II PKSs. This folding pattern is named after the main source of type II PKS namely *Streptomycetes* or S-mode folding. In contrast to F-mode folding

[1] KSβ is also often named chain length factor (CLF)

the lower number in the S-mode folding is odd and the most often found fist-ring folding patterns for type II PKSs are C7-C12 and C9-C14 (Figure 2.3). When the first cyclisation or cyclisations have been formed further ring cyclisations are usually also directed by cyclases. This enables the type II PKSs to steer the otherwise very reactive poly-β-keto chain into a certain structure [14, 25, 39].

Compared to the many different types of type I PKS enzymes and products, type II PKSs and products are less diverse, as the diversity coming from the PKSs themselves are determined by chain length and folding pattern. With that being said, many interesting and important bioactive compounds originate from type II PKSs and actinomycetes continue to be a rich source of novel bioactive compounds, and many existing drugs are fully or partly derived from type II PKSs in actinomycetes [5, 14, 40].

2.1.4 Type III PKSs

The type III PKSs were originally identified as enzymes catalyzing the formation of chalcones and stilbenes in plants. These polyketide synthases utilize the large starter unit *p*-coumaroyl-CoA and through three malonyl-CoA elongations they form chalcones and stilbenes. The first and most intensively studied type III PKS of this type is the chalcone synthase, which is widely distributed in plants. Chalcone is the precursor of flavonoids which have several important roles such as pigment, signalling and defence, and more than 6000 has been described in nature. Naringenin chalcone biosynthesis starts as described by *p*-coumaroyl-CoA being embedded in the active site of the chalcone synthase (CHS). Then three elongations by malonyl-CoA followed by a enzyme specific C1-C6 ring formation yields naringenin chalcone, which is a precursor of flavonoids. The biosynthesis of naringenin chalcone by CHS is depicted in figure 2.9 [12, 41]. Other plant derived type III PKSs also produce products utilizing acetyl-CoA as a starter unit and malonyl-CoA as extender units results in the production of aromatic compounds [42]. Aside from plants, type III PKSs have also been identified and characterized in bacteria [43] and fungi [44].

Type III PKSs function as homodimers of two KS domains. Malonyl-CoA is fed into the active site and the chain is elongated to form an unreduced polyketide chain. During polyketide synthesis the growing chain is momentarily not linked to the type III PKS, but to CoA instead [42]. The mechanism for formation of the polyketide backbone makes type III PKSs leaky and along with the longest

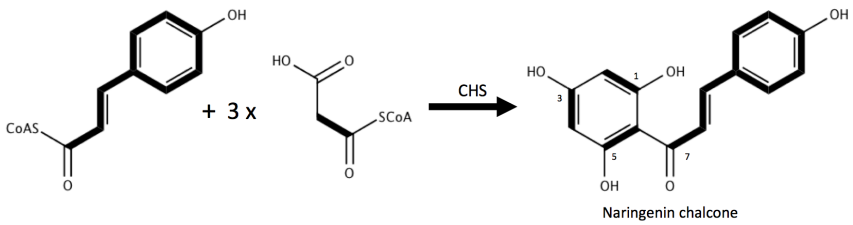


Figure 2.9 The biosynthesis of naringenin chalcone by CHS.

chain produced by the PKS, shorter chains are also formed, likely because the active site allow water to enter, which leads to hydrolysis of premature polyketide chains from the PKS. An example for that is the octaketide synthase OKS from *Aloe aborescens*, which is thought to be responsible for the polyketide production leading to barbaloin. When heterologously expressed the OKS forms hepta- and hexaketide products in easily detectable amounts along with the octaketide [45].

The product from many type III PKSs are poly-β-keto chain, and the ability to selectively fold these are sometimes nested in the PKS itself [41] while only one example exists of a discrete cyclase folding a type III PKS product, being the olivetolic acid cyclase (OAC) involved in biosyntheis of cannabinoids [46]. Many of the investigated type III PKSs have unknown native products as they originate from plants to which only few molecular tools exist. Heterologous expression has therefore been used to investigate these PKSs, which often yields the shunt products formed by the unfolded poly-β-keto chain formed by the type III PKS [12, 45, 47].

2.2 Engineering of PKSs

While the different classes of PKSs and their basic functionality have been investigated for decades, the ability to pick apart a pathway by advances in molecular genetic manipulation and chemical analysis has given us a much better understanding of the mechanisms at work. This has also lead to numerous examples in literature of engineering attempts of these complex enzymes. The following chapter will give examples of some of these attempts to engineer the enzymatic machinery and the resultant products.

2.2.1 Type I mPKS engineering

The principle of colinearity seen for type I mPKSs have lead to scientists to attempt to engineer these large multidomain enzymes. By exchanging modules from one mPKS to another the rationale has been that one should be able to design the resultant polyketide [16]. Early studies of mPKS engineering were producing truncated versions of the mPKSs to verify that shorter products could be obtained [48]. Later studies aimed at investigating the compatibility between different mPKSs. One study did this by keeping the individual modules of the mPKS intact and then swapping domains between two or more mPKSs. Such chimeric mPKSs were made by combining modules by the use of interlinker domains enabling combination of different modules by co-expression of different proteins, which would then interact via these interlinker domains. It was found that the highest production was by domain combinations of different modules from the same mPKSs, ie. module 1 and 3 from DEBS. However, some chimeric mPKSs with modules from different mPKSs produced novel compounds, showing that it to some degree was possible to design compounds from chimeric mPKSs [49, 50]. An alternative approach was to create a chimeric module by fusion of two different modules either from the same mPKS or from two different mPKSs. This was achieved by utilizing homologous recombination (HR) in *Saccharomyces cerevisiae*, which recognized homologous regions between the two modules to form a chimeric module. These chimeric mPKS modules were largely inactive when combining two modules from different mPKSs and active when combining modules from the same system. Therefore, the system can be used to some degree to develop a new module, but the evolutionary distance between the two modules cannot be too large [51].

Instead of combining parts from different mPKSs, another approach has been to selectively mutate domains to be able to design compounds with an alternative oxidation pattern to native compound. This was achieved through selective inactivation by mutation of reducing domains in the mPKS. In this method the important interaction between modules, which had been fine tuned by evolution, was intact by avoiding swapping whole modules [52].

These are some of the attempts to engineer the product of mPKSs and have brought much more insight into the mPKS machinery, but for actual production of polyketides alternative strategies are needed. This is because the engineering attempts of mPKSs have been at the cost of production as the metabolite levels of the engineered mPKSs are often at a tenth or less of the wildtype mPKS [16].

2.2.2 Type I iPKS engineering

The understanding of type I iPKSs have increased substantially over the past decade and with that several attempts have been made to understand and direct product formation of these enzymes. In NR-iPKSs such understanding was gained by creating a minimal type I iPKS, consisting of KS, AT and ACP domains, capable of creating a poly-β-keto chain which then spontaneously folded into the shunt products, which were already known from investigations of minimal type II PKSs [53]. This understanding of the basic domains necessary for chain creation enables the decoupling of chain elongation and folding of the polyketide backbone. By combination of minimal PKSs from one system, and PT and TE domains from other NR-iPKSs it was possible to create novel compounds [23, 54, 55]. Another domain swap strategy has been to make chimera versions of a NR-iPKS in which different proportions of the N-terminus was swapped between two NR-iPKSs. For every domain that swapped the functionality was also changed. This made it possible to change first the starter unit and then chain length by swapping the SAT domain and the KS and AT domain respectively [56]. This ability to swap domains in NR-iPKSs enables design of specific polyketides. An alternative to domain swap has also been selective mutations in PT domain to alter folding specificity [31].

Engineering attempts in PR- or HR-iPKSs are sparse and the catalytic functionality of the domains is still not as well understood as that for NR-iPKSs [57]. One study did domain swapping between two homologous HR-iPKS-NRPS hybrids and found that the KR domain was pivotal in chain length determination and that the C-methylation domains were independent of chain lengths in their functionality [58].

2.2.3 Type II PKS engineering

The nature of type II PKSs allow combinatorial biosynthesis in a more simple fashion than what is needed for type I PKSs. The fact that the system consist of standalone enzymes allow swapping of a whole enzyme instead of having to swap a specific domain. This is simpler as domain swaps require precise determination of domain limitations. However, the substrate specificity and the required protein-protein interactions between the individual enzymes in type II PKSs limits the combinatorial spectrum. One method for development of novel compounds is the combination of a given minimal PKS with a certain chain length co-expressed

with non-native cyclases which has been show to allow control of chain length and folding pattern [59, 60, 61, 62]. The very large number of type II PKSs characterized allows for the utilization of parts from different systems to engineer a compound of interest [34, 63]. This promiscuity of cyclases has also been used in combination with type I [64] and type III PKSs [65, 66] to direct the folding of a nascent poly-β-keto chain.

Another approach to engineer type II PKSs has been to utilize alternative starter units. This has been through derivatising the native aromatic starter unit [67, 68] or by incorporation of a nitrogen containing starter unit into new compounds [69], thereby utilizing the promiscuity of the biosynthetic enzymes. In these studies they also found that the incorporation of alternative starter units were only partially successful as enzymes later in the pathways were sensitive to what starter unit was used [34]. Aside from testing enzyme promiscuity, starter unit specificity engineering to allow fatty acid substrates has also yielded novel compounds by mutational changes in the biosynthetic enzymes [70].

2.2.4 Type III PKS engineering

The simplicity of type III PKSs and the availability of numerous high resolution crystal structures [71, 72, 73, 74] has enabled point mutation engineering of type III PKSs to alter the chain length produced especially by type III PKSs accepting acetyl-CoA as the starter unit [47, 75]. This has enabled the octaketide synthase from *Aloe aborescens* to produce dodecaketide (C_{24}) products, which is the longest product recorded from a type III PKS [76, 77]. As with other PKS systems starter unit promiscuity has also been utilized for derivatization in type III PKS derived products, with focus on different aromatic analogs as starter units for chalcone synthase type III PKS [78] and other complex starter units in the PKS1 from *Huperzia serrata* [79].

Another goal of type III PKS engineering has been on improving the output of the enzymatic machinery. This has been achieved by protein engineering of the 2PS PKS from *Gerbera hybrida* to improve triacetic lactone (TAL) production by 2.5 fold of initial expression [80]. This is in strong contrast to other PKS systems as protein engineering in general has led to reduced overall production of polyketide [16, 34].

2.3 Conclusion and outlook

The drive to discover novel natural products has led the field to also understand the biosynthetic origin of natural products. As exemplified here this knowledge has both been utilized to further understand the enzymatic machinery while it has also enabled design of polyketides which are non-natural. The toolbox to produce any polyketide of choice is still not available, but several techniques exist to begin development of combinatorial enzymatic pipelines catalyzing biosynthesis of new compounds.

References

[1] Nødvig, CS, Nielsen, JB, Kogle, ME, and Mortensen, UH. "A CRISPR-Cas9 system for genetic engineering of filamentous fungi". In: *PLoS One* 10.7 (2015), pp. 1–18.

[2] Chiang, YM, Oakley, CE, Ahuja, M, Entwistle, R, Schultz, A, Chang, SL, Sung, CT, Wang, CCC, and Oakley, BR. "An efficient system for heterologous expression of secondary metabolite genes in *Aspergillus nidulans*". In: *Journal of the American Chemical Society* 135.20 (2013), pp. 7720–7731.

[3] Zhang, W and Tang, Y. "In vitro analysis of type II polyketide synthase". In: *Method Enzymol* 459 (2009), pp. 367–393.

[4] Kildgaard, S, Mansson, M, Dosen, I, Klitgaard, A, Frisvad, JC, Larsen, TO, and Nielsen, KF. "Accurate dereplication of bioactive secondary metabolites from marine-derived fungi by UHPLC-DAD-QTOFMS and a MS/HRMS library". In: *Mar Drugs* 12.6 (2014), pp. 3681–3705.

[5] Dewick, PM. *Medicinal Natural Products - A Biosynthetic Approach, Third Edition*. 3rd. John Wiley & Sons Ltd, 2009.

[6] Rawlings, BJ. "Biosynthesis of fatty acids and related metabolites". In: *Nat Prod Rep* 14.4 (1997), pp. 335–358.

[7] Cox, RJ. "Polyketides, proteins and genes in fungi: Programmed nano-machines begin to reveal their secrets". In: *Org Biomol Chem* 5.13 (2007), pp. 2010–2026.

[8] Fisch, KM. "Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS". In: *RSC Adv* 3.40 (2013), pp. 18228–18247.

[9] Matsuda, Y and Abe, I. "Biosynthesis of fungal meroterpenoids". In: *Nat Prod Rep* 33.1 (2016), pp. 26–53.

[10] Walsh, CT and Tang, Y. *Natural Product Biosynthesis: Chemical Logic and Enzymatic Machinery*. Royal Society of Chemistry, 2017.

[11] Hertweck, C. "The biosynthetic logic of polyketide diversity". In: *Angew Chem Int Ed* 48.26 (2009), pp. 4688–4716.

[12] Yu, D, Xu, F, Zeng, J, and Zhan, J. "Type III polyketide synthases in natural product biosynthesis". In: *IUBMB Life* 64.4 (2012), pp. 285–295.

[13] Xu, W, Qiao, K, and Tang, Y. "Structural analysis of protein-protein interactions in type i polyketide synthases". In: *Crit Rev Biochem Mol Biol* 48.2 (2013), pp. 98–122.

[14] Hertweck, C, Luzhetskyy, A, Rebets, Y, and Bechthold, A. "Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork". In: *Nat Prod Rep* 24.1 (2007), pp. 162–190.

[15] Khosla, C, Tang, Y, Chen, AY, Schnarr, NA, and Cane, DE. "Structure and mechanism of the 6-deoxyerythronolide B synthase". In: *Annu Rev Biochem* 76 (2007), pp. 195–221.

[16] Weissman, KJ. "Genetic engineering of modular PKSs: From combinatorial biosynthesis to synthetic biology". In: *Nat Prod Rep* 33.2 (2016), pp. 203–230.

[17] Hill, AM and Staunton, J. "Type I Modular PKS". In: *Comprehensive Natural Products II - Chemistry and Biology*. Ed. by H Liu and L Mander. Vol. 1. Elsevier, 2010, pp. 385–452.

[18] Wiemann, P, Willmann, A, Straeten, M, Kleigrewe, K, Beyer, M, Humpf, HU, and Tudzynski, B. "Biosynthesis of the red pigment bikaverin in *Fusarium fujikuroi*: genes, their function and regulation". In: *Mol Microbiol* 72.4 (2009), pp. 931–946.

[19] Crawford, JM, Thomas, PM, Scheerer, JR, Vagstad, AL, Kelleher, NL, and Townsend, CA. "Deconstruction of iterative multidomain polyketide synthase function". In: *Science* 320.5873 (2008), pp. 243–246.

[20] Li, Y, Chooi, YH, Sheng, Y, Valentine, JS, and Tang, Y. "Comparative characterization of fungal anthracenone and naphthacenedione biosynthetic pathways reveals an α-hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase". In: *J Am Chem Soc* 133.39 (2011), pp. 15773–15785.

[21] Deng, MR, Guo, J, Li, X, Zhu, CH, Zhu, HH, Deng, MR, Guo, ÁJ, Zhu, CH, Zhu, HH, and Li, X. "Granaticins and their biosynthetic gene cluster from *Streptomyces vietnamensis*: evidence of horizontal gene transfer". In: *Antonie van Leeuwenhoek* 100 (2011), pp. 607–617.

[22] Kim, BC, Lee, JM, Ahn, JS, and Kim, BS. "Cloning, sequencing, and characterization of the pradimicin biosynthetic gene cluster of Actinomadura hibisca P157-2". In: *J Microbiol Biotechnol* 17.5 (2007), pp. 830–839.

[23] Li, Y, Xu, W, and Tang, Y. "Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains". In: *J Biol Chem* 285.30 (2010), pp. 22764–22773.

[24] Liu, L, Zhang, Z, Shao, CL, Wang, JL, Bai, H, and Wang, CY. "Bioinformatical Analysis of the Sequences , Structures and Functions of Fungal Polyketide Synthase Product Template Domains". In: *Sci Rep* 5.10463 (2015), pp. 1–12.

[25] Zhou, H, Li, Y, and Tang, Y. "Cyclization of aromatic polyketides from bacteria and fungi". In: *Nat Prod Rep* 27.6 (2010), pp. 839–868.

[26] Ma, SM, Zhan, J, Watanabe, K, Xie, X, Zhang, W, Wang, C, and Tang, Y. "Enzymatic synthesis of aromatic polyketides using PKS4 from *Gibberella fujikuroi*". In: *J Am Chem Soc* 129.35 (2007), pp. 10642–10643.

[27] Crawford, JM, Dancy, BC, Hill, EA, Udwary, DW, and Townsend, CA. "Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase". In: *Proceedings of the National Academy of Sciences* 103.45 (2006), pp. 16728–16733.

[28] Cox, RJ and Simpson, TJ. "Fungal Type I Polyketides". In: *Comprehensive Natural Products II - Chemistry and Biology*. Ed. by H Liu and L Mander. Vol. 1. Elsevier, 2010, pp. 347–383.

[29] Cacho, RA, Thuss, J, Xu, W, Sanichar, R, Gao, Z, Nguyen, A, Vederas, JC, and Tang, Y. "Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase". In: *J Am Chem Soc* 137.50 (2015), pp. 15688–15691.

[30] Ma, SM, Li, JWH, Choi, JW, Zhou, H, Lee, KM, Moorthie, VA, Xie, X, Kealey, JT, Da Silva, NA, Vederas, JC, and Tang, Y. "Complete reconstitution of a highly reducing iterative polyketide synthase". In: *Science* 326.5952 (2009), pp. 589–592.

[31] Xu, Y, Zhou, T, Zhou, Z, Su, S, Roberts, SA, Montfort, WR, Zeng, J, Chen, M, Zhang, W, Lin, M, Zhan, J, and Molnar, I. "Rational reprogramming of fungal polyketide first-ring cyclization". In: *Proc Nat Acad Sci* 110.14 (2013), pp. 5398–5403.

[32] Mao, XM, Zhan, ZJ, Grayson, MN, Tang, MC, Xu, W, Li, YQ, Yin, WB, Lin, HC, Chooi, YH, Houk, K, and Tang, Y. "Efficient biosynthesis of fungal polyketides containing the dioxabicyclo-octane ring system". In: *J Am Chem Soc* 137.37 (2015), pp. 11904–11907.

[33] Lin, TS, Chiang, YM, and Wang, CCC. "Biosynthetic pathway of the reduced polyketide product citreoviridin in *Aspergillus terreus* var. *aureus* revealed by heterologous expression in *Aspergillus nidulans*". In: *Org Lett* 18.6 (2016), pp. 1366–1369.

[34] Rohr, J and Hertweck, C. "Type II PKS". In: *Comprehensive Natural Products II - Chemistry and Biology*. Ed. by L Mander and HW Liu. Vol. 1. Elsevier, 2010, pp. 227–303.

[35] Das, A and Khosla, C. "Biosynthesis of aromatic polyketides in bacteria". In: *Acc Chem Res* 42.5 (2009), pp. 631–639.

[36] Reeves, CD. "The enzymology of combinatorial biosynthesis". In: *Crit Rev Biotechnol* 23.2 (2003), pp. 95–147.

[37] Carreras, CW, Rembert, P, and Khosla, C. "Efficient synthesis of aromatic polyketides in vitro by the actinorhodin polyketide synthase". In: *J Am Chem Soc* 118.21 (1996), pp. 5158–5159.

[38] Chen, A, Re, RN, and Burkhart, MD. "Type II fatty acid and polyketide synthases: deciphering protein–protein and protein–substrate interactions". In: *Nat Prod Rep* 35.10 (2018), pp. 1029–1045.

[39] Thomas, R. "A biosynthetic classification of fungal and streptomycete fused-ring aromatic polyketides". In: *ChemBioChem* 2.9 (2001), pp. 613–627.

[40] van Keulen, G and Dyson, PJ. "Production of specialized metabolites by *Streptomyces coelicolor* A3 (2)". In: *Adv App Microbiol* 89 (2014), pp. 217–266.

[41] Austin, MB and Noel, JP. "The chalcone synthase superfamily of type III polyketide synthases". In: *Nat Prod Rep* 20.1 (2003), pp. 79–110.

[42] Abe, I and Morita, H. "Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases". In: *Nat Prod Rep* 27.6 (2010), pp. 809–838.

[43] Li, Y and Müller, R. "Non-modular polyketide synthases in myxobacteria". In: *Phytochem* 70.15–16 (2009), pp. 1850–1857.

[44] Hashimoto, M, Nonaka, T, and Fujii, I. "Fungal type III polyketide synthases". In: *Nat Prod Rep* 31.10 (2014), pp. 1306–1317.

[45] Mizuuchi, Y, Shi, SP, Wanibuchi, K, Kojima, A, Morita, H, Noguchi, H, and Abe, I. "Novel type III polyketide synthases from *Aloe arborescens*". In: *FEBS J* 276.8 (2009), pp. 2391–2401.

[46] Gagne, SJ, Stout, JM, Liu, E, Boubakir, Z, Clark, SM, and Page, JE. "Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides". In: *Proc Nat Acad Sci* 109.31 (2012), pp. 12811–12816.

[47] Abe, I, Oguro, S, Utsumi, Y, Sano, Y, and Noguchi, H. "Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase". In: *J Am Chem Soc* 127.36 (2005), pp. 12709–12716.

[48] Wiesmann, KEH, Cortés, J, Brown, MJB, Cutter, AL, Staunton, J, and Leadlay, PF. "Polyketide synthesis in vitro on a modular polyketide synthase". In: *Cell Chem Biol* 2.9 (1995), pp. 583–589.

[49] Menzella, HG, Reid, R, Carney, JR, Chandran, SS, Reisinger, SJ, Patel, KG, Hopwood, DA, and Santi, DV. "Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes". In: *Nat Biotechnol* 23.9 (2005), pp. 1171–1176.

[50] Menzella, HG, Carney, JR, and Santi, DV. "Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases". In: *Cell Chem Biol* 14.2 (2007), pp. 143–151.

[51] Chemler, JA, Tripathi, A, Hansen, DA, O'Neil-Johnson, M, Williams, RB, Starks, C, Park, SR, and Sherman, DH. "Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination". In: *J Am Chem Soc* 137.33 (2015), pp. 10603–10609.

[52] Kushnir, S, Sundermann, U, Yahiaoui, S, Brockmeyer, A, Janning, P, and Schulz, F. "Minimally invasive mutagenesis gives rise to a biosynthetic polyketide library". In: *Angew Chem Int Ed* 51.42 (2012), pp. 10664–10669.

[53] Zhang, W, Li, Y, and Tang, Y. "Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*". In: *Proc Nat Acad Sci* 105.52 (2008), pp. 20683–20688.

[54] Newman, AG, Vagstad, AL, Storm, PA, and Townsend, CA. "Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming". In: *J Am Chem Soc* 136.20 (2014), pp. 7348–7362.

[55] Vagstad, AL, Newman, AG, Storm, PA, Belecki, K, Crawford, JM, and Townsend, CA. "Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products". In: *Angew Chem Int Ed* 52.6 (2013), pp. 1718–1721.

[56] Liu, T, Sanchez, JF, Chiang, YM, Oakley, BR, and Wang, CC. "Rational domain swaps reveal insights about chain length control by ketosynthase domains in fungal nonreducing polyketide synthases". In: *Org Lett* 16.6 (2014), pp. 1676–1679.

[57] Simpson, TJ. "Fungal polyketide biosynthesis-A personal perspective". In: *Nat Prod Rep* 31.10 (2014), pp. 1247–1252.

[58] Fisch, KM, Bakeer, W, Yakasai, AA, Song, Z, Pedrick, J, Wasil, Z, Bailey, AM, Lazarus, CM, Simpson, TJ, and Cox, RJ. "Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite". In: *J Am Chem Soc* 133.41 (2011), pp. 16635–16641.

[59] Tang, Y, Lee, TS, and Khosla, C. "Engineered biosynthesis of regioselectively modified aromatic polyketides using bimodular polyketide synthases". In: *PLoS Biol* 2.2 (2004), pp. 227–238.

[60] Metsä-Ketelä, M, Palmu, K, Kunnari, T, Ylihonko, K, and Mäntsälä, P. "Engineering anthracycline biosynthesis toward angucyclines". In: *Antimicrob Agents Chemother* 47.4 (2003), pp. 1291–1296.

[61] McDaniel, R, Ebert-Khosla, S, Hopwood, DA, and Khosla, C. "Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits". In: *Nature* 375.6532 (1995), pp. 549–554.

[62] Rawlings, BJ. "Biosynthesis of polyketides (other than actinomycete macrolides)". In: *Nat Prod Rep* 16.4 (1999), pp. 425–484.

[63] Kim, J and Yi, GS. "PKMiner: A database for exploring type II polyketide synthases". In: *BMC Microbiol* 12 (2012), pp. 1–12.

[64] Ma, S, Zhan, J, Xie, X, Watanabe, K, Tang, Y, and Zhang, W. "Redirecting the cyclization steps of fungal polyketide synthase". In: *J Am Chem Soc* 130.1 (2008), pp. 38–39.

[65] Andersen-Ranberg, J, Kongstad, KT, Nafisi, M, Staerk, D, Okkels, FT, Mortensen, UH, Lindberg Møller, B, Frandsen, RJN, and Kannangara, R. "Synthesis of C-Glucosylated Octaketide Anthraquinones in *Nicotiana benthamiana* by Using a Multispecies-Based Biosynthetic Pathway". In: *ChemBioChem* 18.19 (2017), pp. 1893–1897.

[66] Frandsen, RJ, Khorsand-Jamal, P, Kongstad, KT, Nafisi, M, Kannangara, RM, Staerk, D, Okkels, FT, Binderup, K, Madsen, B, Møller, BL, Thrane, U, and Mortensen, UH. "Heterologous production of the widely used natural food colorant carminic acid in *Aspergillus nidulans*". In: *Sci Rep* 8.12853 (2018), pp. 1–10.

[67] Kalaitzis, JA, Izumikawa, M, Xiang, L, Hertweck, C, and Moore, BS. "Mutasynthesis of enterocin and wailupemycin analogues". In: *J Am Chem Soc* 125.31 (2003), pp. 9290–9291.

[68] Izumikawa, M, Cheng, Q, and Moore, BS. "Priming type II polyketide synthases via a type II nonribosomal peptide synthetase mechanism". In: *J Am Chem Soc* 128.5 (2006), pp. 1428–1429.

[69] Zhang, W, Ames, BD, Tsai, SC, and Tang, Y. "Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase". In: *Appl Environ Microbiol* 72.4 (2006), pp. 2573–2580.

[70] Nicholson, TP, Winfield, C, Westcott, J, Crosby, J, Simpson, TJ, and Cox, RJ. "First in vitro directed biosynthesis of new compounds by a minimal type II polyketide synthase: Evidence for the mechanism of chain length determination". In: *Chem Commun* 9.6 (2003), pp. 686–687.

[71] Morita, H, Kondo, S, Oguro, S, Noguchi, H, Sugio, S, Abe, I, and Kohno, T. "Structural insight into chain-length control and product specificity of pentaketide chromone synthase from Aloe arborescens". In: *Cell Chem Biol* 14.4 (2007), pp. 359–369.

[72] Austin, MB, Bowman, ME, Ferrer, JL, Schröder, J, and Noel, JP. "An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases". In: *Cell Chem Biol* 11.9 (2004), pp. 1179–1194.

[73] Morita, H, Shimokawa, Y, Tanio, M, Kato, R, Noguchi, H, Sugio, S, Kohno, T, and Abe, I. "A structure-based mechanism for benzalacetone synthase from Rheum palmatum". In: *Proc Nat Acad Sci* 107.2 (2010), pp. 669–673.

[74] Ferrer, JL, Jez, JM, Bowman, ME, Dixon, RA, and Noel, JP. "Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis". In: *Nat Struct Mol Biol* 6.8 (1999), pp. 775–784.

[75] Abe, I, Utsumi, Y, Oguro, S, Morita, H, Sano, Y, and Noguchi, H. "A plant type III polyketide synthase that produces pentaketide chromone". In: *J Am Chem Soc* 127.5 (2005), pp. 1362–1363.

[76] Wakimoto, T, Morita, H, and Abe, I. “Engineering of plant type III polyketide synthases”. In: *Method Enzymol* 515 (2012), pp. 337–358.

[77] Wanibuchi, K, Morita, H, Noguchi, H, and Abe, I. “Enzymatic formation of an aromatic dodecaketide by engineered plant polyketide synthase”. In: *Bioorg Med Chem Lett* 21.7 (2011), pp. 2083–2086.

[78] Abe, I, Utsumi, Y, Oguro, S, and Noguchi, H. “The first plant type III polyketide synthase that catalyzes formation of aromatic heptaketide”. In: *FEBS lett* 562.1-3 (2004), pp. 171–176.

[79] Morita, H, Yamashita, M, Shi, SP, Wakimoto, T, Kondo, S, Kato, R, Sugio, S, Kohno, T, and Abe, I. “Synthesis of unnatural alkaloid scaffolds by exploiting plant polyketide synthase”. In: *Proc Nat Acad Sci* 108.33 (2011), pp. 13504–13509.

[80] Vickery, CR, Cardenas, J, Bowman, ME, Burkart, MD, Da Silva, NA, and Noel, JP. “A coupled *in vitro/in vivo* approach for engineering a heterologous type III PKS to enhance polyketide biosynthesis in *Saccharomyces cerevisiae*”. In: *Biotechnol Bioeng* 115.6 (2018), pp. 1394–1402.

3 Manuscript I – Steering polyketide production in yeast by combinatorial expression of enzymes from type II and III polyketide synthase systems

Kromphardt KJK[a], Skovbjerg CAS[a], Coumou HC[a,b], Larsen TO[a] & Frandsen RJN[a]

This manuscript is currently in preparation.

[a] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark; [b] currently Novozymes, Bagsværd, Denmark

3.1 Abstract

The discovery and characterization of new polyketides has for a long time been driven by the fact that many polyketides possess desirable bioactivities. Besides the discovery of novel polyketides, much effort has also been put into understanding the enzymes, named polyketide synthases (PKSs), that synthesize these complex compounds. This has also brought about several attempts of altering the natural products produced by the individual PKSs by protein engineering. Especially, the PKSs that produce producing non-reduced polyketides have received attention as this subgroup of polyketides having cyclic and aromatic structures, which often display relevant biological activities. In the current study we set out to provide proof-of-concept for a programmable microbial platform for the production of cyclic and aromatic polyketides, which would allow control of the chain length and folding pattern of the produced chain. This was achieved through co-expression of different type III PKSs and type II polyketide cyclases catalyzing different folding patterns, creating an artificial PKS systems. Expression of five plant derived type III PKSs in *S. cerevisiae* resulted in the production of their previously described products, varying in length from C_6 to C_{16}. Co-expression of the octaketide (C_{16}) producing type III PKS with two different polyketide cyclases from actinomycete bacteria, allowed for directing different folding patterns. The cyclases were both found to be able to direct the folding of the octaketide backbone, which was produced by the type III PKS. This proof-of-concept study lays the foundation for

further development of the platform, as it shows that it is possible to rationally program this artificial PKS system.

Keywords: Yeast, Metabolic engineering, Polyketide synthases

3.2 Introduction

Many pharmacological compounds in clinical use are completely or partially derived from natural sources. This has brought much interest into the discovery of new compounds from natural sources as nature has shown to be an excellent source of such compounds [1, 2]. Many of the pharmacologically active compounds are of polyketide origin, and much effort has been put into understanding the biosythesis of polyketides [3, 4, 5, 6]. Polyketides are biosynthesized by polyketide synthases (PKSs) which are divided into three types (I, II and III) based on their tertiary structure and mode of catalysis [7].

Much effort has been given to understanding the synthesis of the non-reduced polyketides, as this subgroup is rich in bioactive compounds [7, 8]. Synthesis mechanism of non-reduced polyketides can be split into two main steps. The first being the synthesis of the non-reduced polyketide chain and the second being folding of the chain in a given pattern to form a cyclic and aromatic structure [7]. Folding of the polyketide chain can either be carried out by the same enzyme producing the chain, which is observed in type I iterative PKSs [8], or by an trans acting enzyme, which is observed in type II PKS systems [5]. The two given examples are named F-mode and S-mode respectively referring to their fungal or streptomycete origin (Figure 3.1). The known common F-mode folding patterns referring to the first carbon-carbon bond being formed are C2-C7, C4-C9 or C6-C11, while the known common S-mode folding patterns are C7-C12 or C9-C14 first ring folding [9]. Even though type III PKSs often produce cyclic and aromatic polyketides only a single cyclase has so far been described to fold a type III PKS product in a natural system [10] and the general folding mechanisms for type III PKS systems remain unknown.

The growing understanding of the biosynthesis of polyketides has over the years lead to many attempts to engineer PKSs and thus alter the natural product of many different PKSs. This has been achieved through domain engineering [11], domain swapping [12, 13] and combinatorial libraries with components from different natural pathways [14, 15] to name a few. The mentioned examples have been engineering of PKSs within the same system while product engineering by

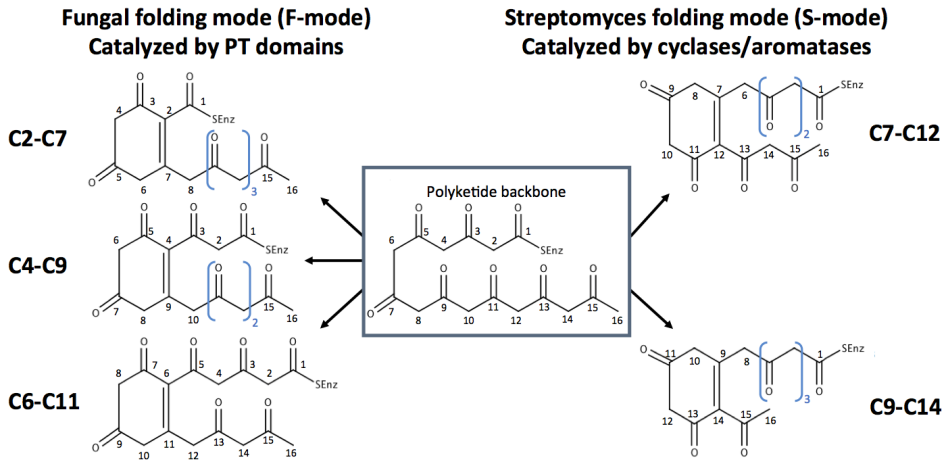


Figure 3.1 The principle behind the F- and S-mode first ring folding in natural products.

combination of components from different PKS systems has also been made [16, 17]. Generally, applicable methods for rational reprogramming of PKS systems to design a specific compound of interest has long been the goal of the research field. It has, however, proven almost impossible to device general strategies as the engineered PKSs typically loose the majority of their catalytic activity due to incompatibilities between the different heterologous PKS elements that are combined in the artificial systems.

In this study, we combine elements from two of the natural PKS types in an attempt to bypass the protein-protein interaction challenges faced when combining elements from different PKS belonging to the same PKS type. The aim of this study was first to prove that a diverse set of plant derived type III PKSs with a diversity of products could be functionally expressed in *Saccharomyces cerevisiae*. Secondly, co-expression of a series of actinomycete derived cyclases from type II PKS systems with the type III PKSs and test if they would be able to direct the folding pattern of the produced polyketides. Type II cyclases and type III PKSs have previously been shown to be able to produce folded products when expressed in tobacco and *Aspergillus nidulans* [17, 18]. However, heterologous expression in *S. cerevisiae* has not been tested, though it would provide a cleaner metabolic background for metabolic engineering, and make the system more relevant for industrial scale fermentation in connection with production of beverages [19] and biopharmaceuticals [20].

3.3 Materials and methods

3.3.1 Strains, vectors, media and synthetic genes

S. cerevisiae strains constructed in this study were derived from the CEN.PK102-5B or CEN.PK111-61A [21]. An overview of all strains can be seen in table 3.1. Episomal 2μ vector pBOSAL1-URA [22] as well as integrative vectors [23] were used for expression of genes in *S. cerevisiae*. Yeast strains were maintained on yeast extract peptone dextrose (YPD) medium (10 g/L Bacto yeast extract, 20 g/L Bacto Peptone, 20 g/L glucose), except for strains carrying a plasmid or during transformation steps. For transformation or plasmid maintenance yeast strains were grown on synthetic complete medium lacking leucine (SC-leu) or uracil (SC-ura) [24] for selection of the 2μ plasmid carrying the LEU2 gene from *Kluyveromyces lactis* or integrative transformation cassette carrying the URA3 gene from *Kluyveromyces lactis*. *Escherichia coli* strain DH5α was used for all vector assemblies and propagation [25]. *E. coli* strains were all grown in or on lysogeny broth medium supplemented with 100 µg/L ampicillin (LB-Amp). The polymerase *Pfu*X7 [26] was used for all PCR reactions. All primers (Integrated DNA technologies, Belgium) used in this study are listed in appendix (Table S3.1). Synthetic genes Gh2PS (P48391) [27], AaPCS (Q58VP7) [28], DluHKS (ABQ59603) [29], AaPKS3 (C4MBZ5) [30], AaOKS (Q3L7F5) [31], Gra-orf4 (GU233672) [32] and PdmD (EF151801) [33] were ordered from and codon optimized for *S. cerevisiae* by Genscript (NJ, USA).

3.3.2 Vector and strain construction

Vectors were assembled by Uracil-Specific Excision Reagent (USER) fusion as described in Jensen *et al.* (2014) [23]. An overview of all vectors can be found in table 3.1. The yeast promoters P_{TEF1} and P_{PGK1} were amplified from yeast genomic DNA from CEN.PK117-7D [21]. All plasmids were verified by restriction analysis and sequencing by Eurofins MGW Operon (Ebersberg, Germany). *S. cerevisiae* transformations were performed via the LiAc/SS-DNA/PEG method [34]. Prior to transformation of integrative vectors these were linearized by restriction as described in Jensen *et al.* (2014) [23]. Transformation of integration of fragments were verified by diagnostic colony PCR. Diploid strains were constructed from the the haploid strains by mating on YPD for 6-8 hours, after which diploids were

Table 3.1 List of vectors and *S. cerevisiae* strains used and constructed in this study.

Vector name	Description	Source
pBOSAL1	2µ *Kl.LEU2*	[22]
pBOSAL1-Gh2PS	2µ P_{PGK1}-*Gh2PS*-T_{ADH1}-*Kl.LEU2*	This study
pBOSAL1-AaPCS	2µ P_{PGK1}-*AaPCS*-T_{ADH1}-*Kl.LEU2*	This study
pBOSAL1-DluHKS	2µ P_{PGK1}-*DluHKS*-T_{ADH1}-*Kl.LEU2*	This study
pBOSAL1-AaPKS3	2µ P_{PGK1}-*AaPKS3*-T_{ADH1}-*Kl.LEU2*	This study
pBOSAL1-AaOKS	2µ P_{PGK1}-*AaOKS*-T_{ADH1}-*Kl.LEU2*	This study
pCfB389	Int. XI-2 *Kl.URA3*	[23]
pCfB257	Int. X-3 *Kl.LEU2*	[23]
pCfB389-PdmD	Int. XI-2 P_{TEF1}-*PdmD*-T_{CYC1}-*Kl.URA3*	This study
pCfB389-Gra-orf4	Int. XI-2 P_{TEF1}-*Gra-orf4*-T_{CYC1}-*Kl.URA3*	This study

Strain name	Description	Source
CEN.PK113-7D	*MATa* MAL2-8c SUC2	P. Kötter
CEN.PK102-5B	*MATa MAL2-8c SUC2 ura3-52 his3Δ1 leu2-3/112*	P. Kötter
CEN.PK111-61A	*MATα MAL2-8c SUC2 ura3-52 his3Δ1 leu2-3/112*	P. Kötter
Ant-SC117	*CEN.PK102-5B pBOSAL1-LEU2*	This study
Ant-SC115	*CEN.PK102-5B pBOSAL1-Gh2PS*	This study
Ant-SC136	*CEN.PK102-5B pBOSAL1-AaPCS*	This study
Ant-SC125	*CEN.PK102-5B pBOSAL1-DluHKS*	This study
Ant-SC126	*CEN.PK102-5B pBOSAL1-AaPKS3*	This study
Ant-SC116	*CEN.PK102-5B pBOSAL1-AaOKS*	This study
Ant-SC090	*CEN.PK111-61A XI-2::*P_{TEF1}-*pdmD*-T_{CYC1}-*Kl.URA3*	This study
Ant-SC091	*CEN.PK111-61A XI-2::*P_{TEF1}-*gra-orf4*-T_{CYC1}-*Kl.URA3*	This study
Ant-SC110	*CEN.PK111-61A XI-2::Kl.URA3*	This study
Ant-SC111	*CEN.PK102-5B X-3::Kl.LEU2*	This study
Ant-SC127	Ant-SC116 x Ant-SC090 (diploid)	This study
Ant-SC128	Ant-SC116 x Ant-SC091 (diploid)	This study
Ant-SC132	Ant-SC116 x Ant-SC110 (diploid)	This study
Ant-SC207	Ant-SC111 x Ant-SC090 (diploid)	This study
Ant-SC209	Ant-SC111 x Ant-SC091 (diploid)	This study

selected for on SC lacking uracil and leucine (Sc-ura/leu).

3.3.3 Cultivation, chemical extraction and chemical analysis

To prepare strains for chemical extraction these were grown for 5 days at 30°C in 500 mL Erlenmeyer flasks with baffles containing 50mL liquid SC-leu at 200 rpm. All flask were inoculated from an overnight culture to have OD_{600} = 0.1 to initiate the cultivation. OD_{600} measurements were done on a Shimadzu UV-1800 spectrophotometer (Japan).

Chemical extractions were done by separating the biomass from the culture liquid via centrifugation (8000 xg for 5 minutes) in a 50mL falcon tube (Subsequently named tube A). The culture liquid was decanted into another 50mL falcon tube (Subsequently named tube B). To lyse the yeast cells 25mL of extraction liquid (ethylacetate with 1% v/v formic acid) was added to the pellet in tube A and subjected to ultrasound for 1 hour. Afterwards, 25mL of the culture liquid from tube B was decanted into the extraction liquid and yeast cell mixture. After vigorous shaking the culture liquid was removed with a glass Pasteur pipette from tube

A. Tube A was centrifuged (8000 xg for 5 minues) and the extraction liquid was decanted into the remaining culture liquid tube B. After vigorous shaking with the remaining culture liquid the culture liquid was removed with a glass Pasteur pipette. The extraction liquid was evaporated to dryness under a gentle stream of N_2 at room temperature. The residue was redissolved in 300µL methanol and exposed to ultrasound for 10 seconds and transferred to a 2 mL Eppendorph tube. The Eppendorph tube was centrifuged (18000 xg for 5 minutes) and 150µL of methanol was transferred to a HPLC-vial.

Chemical analysis was done by Ultra high performance-diode array detection-time of flight mass spectrometry (UHPLC-DAD-TOFMS) on a Bruker maXis G3 QTOF mass spectrometer (Bruker Daltronics, Bremen, Germany) which used an electrospray ionization source (ESI). Sample separation was done on a Dionex Ultima 3000 UHPLC-DAD (Dionex, Sunnyvale, CA, USA) with a Kinetex C_{18} column (150 x 2.1 mm, 2.6 μm (Phenomenex Inc., Torrance, CA, USA)) which was kept at 40°C and a mobile phase of a gradient mixture of water (A) and acetonitrile (B) both containing 20mM formic acid. The gradient profile started at 10% B and increased to a 100% B in 10 min, held there for 3 min and then returned to 10% in 0.1 min and kept there for 1.5 min with a flowrate of 400µL/min. Mass spectra were recorded for a mass range of m/z 75-1250 in negative MS mode and UV spectra was collected from 200-700 nm.

3.3.4 Large scale cultivation, purification of compounds and NMR structure elucidation

To produce sufficient biomass for purification of compounds for NMR based structure elucidation a larger scale cultivation was setup. The medium was synthetic minimal (SM) [35] supplemented 125 mg/L histidine and 150 mg/L uracil [36]. The strain Ant-SC116 was chosen for purification of selected compounds. The larger scale cultivation was carried out in 2L BioStat B bioreactor (B. Braun Biotech International, Germany). The cultivation was initiated by inoculation to a start OD_{600} = 0.01. Throughout the cultivation the pH was kept at 5 (adjusted with 2M NaOH and 2M H_2SO_4), the temperature was 30°C, the aeriation was 2L/min and agitation was at 600 rpm. Two reactors were run resulting in 4L of total culture.

Culture liquid and yeast cells were separated by storing the culture liquid at 5°C for 10 days. Yeast cells precipitated at the bottom and the supernatant was

decanted. The supernatant was pH adjusted to 4 and extracted with two volumes of ethylacetate. The precipitated yeast cells were redissolved in 30 mL methanol and subjected to ultrasound for 1h. After a centrifugation step (8000 xg, 5 min), the methanol fraction was added to the ethylacetate and evaporated to dryness in a rotavapor. The raw extract (1.2 g) was adsorbed onto Diol material and dried before packing onto a 25g SNAP column (Biotage, Uppsala, Sweden) with Diol material. The extract was fractionated on an Isolera One flash purification system (Biotage) via seven steps of heptane (Hep), dichloromethane (DCM), ethylacetate (EtOAc) and methanol (MeOH), with each step having a volume of three column volumes. The steps were as follows: 100% Hep, 50/50% Hep/DCM, 100% DCM, 50/50% DCM/EtOAc, 100% EtOAc, 50/50% EtOAc/MeOH and 100% MeOH. The methanol fraction (212.4 mg) contained the compounds of interest and were subjected to further fractionation on a semi-preparative HPLC on a Waters 600 Controller with a 996 photodiode array detector (Waters, MA, USA). The solid phase was a Phenomenex Luna II C18 column (250x10 mm, 5µm (Phenomenex, CA, USA)) and a mobile phase consisting of a mixture of water (A) and acetonitrile (B) both containing 50ppm triflouroacetic acid. The gradient profile started at 25% B, increased to a 30% B in 10 min, increased to 100% in 1 min, held there for 2 min and then returned to 25% in 1 min and kept there for 2 min with a flowrate of 5 mL/min. The resultant pure compounds were 1 mg of SEK4, 0.5 mg of SEK4b, 1 mg dSEK4 and 1.1 mg of dSEK4b.

The isolated compounds were analyzed by NMR on a Bruker AVANCE 800 MHz (Bruker Daltronics, MA, USA) with a 5mm prodigy cryoprobe using standard pulse sequences (^{1}H-NMR, ^{13}C-NMR, HSQC and HMBC). All spectra were recorded in DMSO-d_6. The signals of residual solvent protons and carbons were used for reference (δ_H 2.50 and δ_C 39.5 ppm for DMSO-d_6). All NMR data were analyzed on Bruker topspin software (Bruker Daltronics, MA, USA). In general NMR signals were not very strong and some NMR signals were missing in our data when comparing to previous isolation of the compounds in literature. The missing or altered signals are mentioned below for each compound. All recorded spectra can be found in appendix (Figure S3.2 to S3.5).

SEK4 (1 mg): UV: $\lambda_{max} = 280$ nm. ESI-MS: m/z 317.0667 [M-H]$^-$. ^{1}H-NMR δ_H ppm: 1.56 (3H, s); 2.54 (1H, d, J = 15.9 Hz); 2.92 (1H, d, J = 15.9 Hz); 4.07 (1H, d, J = 15.9 Hz); 4.15 (1H, d, J = 15.9 Hz); 5.14 (1H, s); 5.61 (1H, s); 6.26 (1H, d, J = 2.4 Hz); 6.32 (1H, d, J = 2.4 Hz); 6.88 (1OH, s); 10.54 (1OH, s);

11.06 (1H, s-broad). ^{13}C-NMR δ_H ppm: 27.95; 38.05; 49.81; 88.55; 101.12; 103.36; 111.83; 113.48; 139.21; 162.40; 163.45; 164.42; 191.56. Apart from three missing carbon signals (99.6; 165.4; 170.5) the recorded NMR spectra and assignments were in agreement with published data [15].

SEK4b (0.5 mg): UV: λ_{max} = 280 nm. ESI-MS: m/z 317.0668 [M-H]$^-$. ^{1}H-NMR δ_H ppm: 2.46 (3H, s); 2.65 (1H, s); 2.91 (1H, d, J = 13.8); 2.98 (1H, d, J = 16.6 Hz); 2.99 (1H, d, J = 13.8); 6.17 (1H, d J = 2.19 Hz); 6.26 (1H, d, J = 1.79 Hz); 6.52 (1H, s); 7.1 (1OH, s); 10.37 (1OH, s); 11.91 (1OH, s-broad). ^{13}C-NMR δ_H ppm: 22.97; 44.73; 47.96. The proton at 5.32 and carbons at 88.9, 100.2, 101.6, 104.6, 111.9, 112.7, 142.5, 160.4, 161.4, 162.8, 164.0, 172.2, and 192.3 were not detected in our dataset. Also, the proton at 2.65 was a singlet in our data, but was supposed to create correlate to the proton at 2.98 which is a double with a J = 16.6 Hz, as they are attached to the same carbon 47.96, which is also revealed in the HSQC spectrum. The absence of this doublet was likely due to the low signal intensity and thus unclear peak assignment. Remaining detected signals were in agreement with previously published data [37].

dSEK4 (1 mg): UV: λ_{max} = 293 nm. ESI-MS: m/z 299.0560 [M-H]$^-$. ^{1}H-NMR δ_H ppm: 2.30 (3H, s); 4.36 (2H, s); 5.11 (1H, s); 5.54 (1H; s); 6.02 (1H, s); 6.73 (1H, d, J = 2.04 Hz), 6.77 (1H, d, J = 2.04 Hz); 10.86 (1OH, s); 11.67 (1OH, s). ^{13}C-NMR δ_H ppm: 19.83; 37.74; 88.41; 100.70; 102.53; 111.11; 114.16; 118.05;.138.35; 159.72; 161.72; 164.46; 164.83; 165.8; 178.15. Protons at 5.11 and 5.54 were singlets in our dataset and the carbon at 170.4 ppm was not detected. These protons should be doublets due to aromatic interactions which was not detected in our dataset. Remaining NMR signals were in agreement with published data [38].

dSEK4b (1.1 mg): UV: λ_{max} = 292 nm. ESI-MS: m/z 299.0564 [M-H]$^-$. ^{1}H-NMR δ_H ppm: 2.65 (3H, s); 3.85 (2H, s); 5.18 (1H, s); 6.09 (2H, s); 6.61 (1H, d, J = 2.16); 6.64 (1H, d, J = 2.13); 10.67 (1OH, s); 11.98 (1OH, s-broad). ^{13}C-NMR δ_H ppm: 22.83; 37.51; 89.15; 101.03; 103.33; 112.59; 114.83; 117.29; 142.19; 159.62; 161.56; 161.70; 164.15; 178.51. The carbons at 160.8 and 172.2 were not detected in our data but all other signals and assignments were in agreement with published data [38].

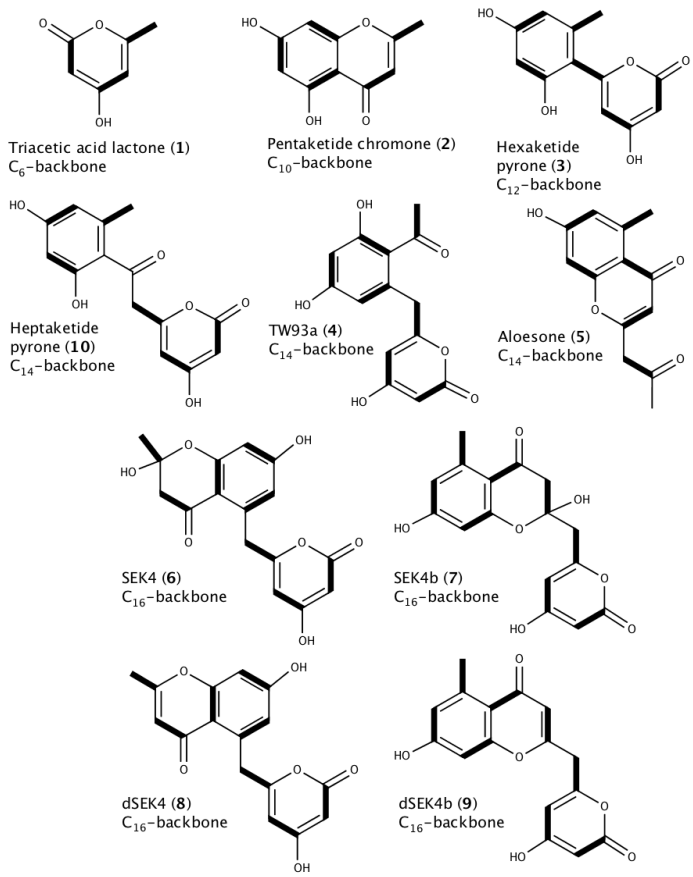


Figure 3.2 Compounds produced by type III PKSs used in this study. The highlighted bonds represent C_2 ketide units added in polyketide synthesis.

Table 3.2 Expressed type III PKSs, their origin and their reported products. Structures of products can be found in figure 3.2.

Name	Organism	Expected primary product(s)	Ref.
Gh2PS	*G. hybrida*	triacetic acid lactone (**1**)	[27]
AaPCS	*A. aborescens*	5,7-dihydroxy-2-methylchromone (**2**)	[28]
DluHKS	*D. lusitanicum*	6-(2',4'-dihydroxy-6'-methylphenyl)-4-hydroxy-2-pyrone (**3**)	[29]
AaPKS3	*A. aborescens*	TW93a (**4**) and aloesone (**5**)	[30]
AaOKS	*A. aborescens*	SEK4 (**6**), SEK4b (**7**), dSEK4 (**8**) and dSEK4b (**9**)	[31]

3.4 Results and discussion

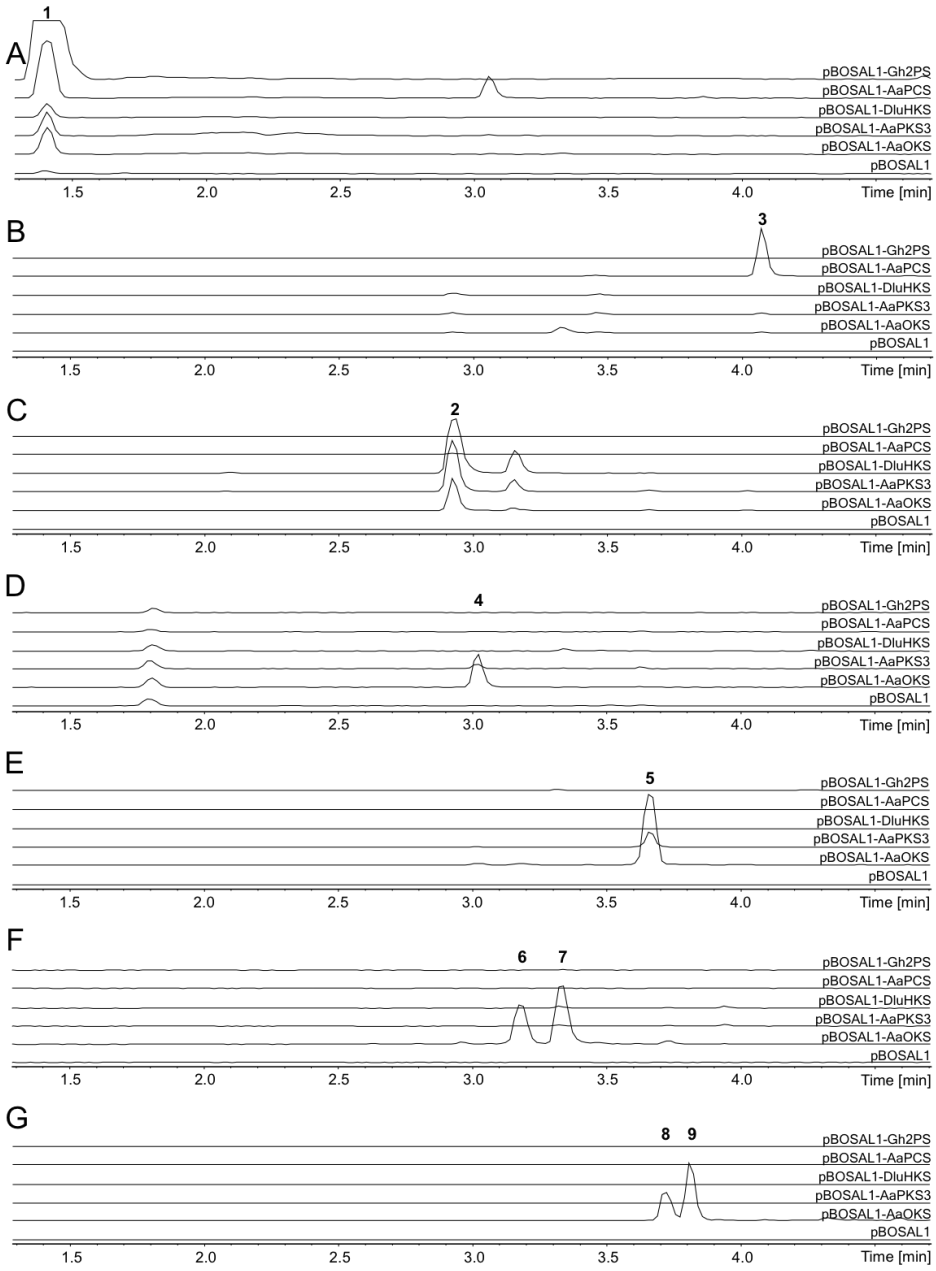
3.4.1 Expression of type III PKSs in *S. cerevisiae*

The initial steps toward building the envisioned platform was to test if the chosen type III PKSs would be functionally expressed in *S. cerevisiae*. To make the product spectrum as wide as possible the type III PKSs were selected on the basis of producing polyketides of different lengths (C_6 to C_{16}) along with using either malonyl-CoA or acetyl-CoA as starter units. The selected type III PKSs were the 2-pyrone synthase from *Gerbera hybrida* (Gh2PS) which produces triacetic acid lactone (**1**) [27] and has previously been functionally expressed in *S. cerevisiae* [39], the hexaketide synthase from *Drosophyllum lusitanicum* (DluHKS) which produces the hexaketide 6-(2',4'-dihydroxy-6'-methylphenyl)-4-hydroxy-2-pyrone (**3**) [29] and the pentaketide synthase (AaPCS), PKS3 (AaPKS3) and octaketide synthase (AaOKS) all from *Aloe aborescens* producing 5,7-dihydroxy-2-methylchromone (**2**) [28]; TW93a **(4)** and aloesone (**5**) [30]; and SEK4 (**6**), SEK4b (**7**), dSEK4b (**8**), and dSEK4 (**9**) [18, 31] respectively as their primary products. An overview of the expressed type III PKSs and their reported primary products can be found in table 3.2 and structures of the reported compounds can be found in figure 3.2.

LC-MS based identification of compounds

To putatively identify the compounds expected to be produced by the heterologously expressed PKSs we used the exact masses of the expected compounds to create EICs. Furthermore, UV spectra and relative retention time of the produced compounds were compared to other polyketides from previous studies when available. UV-spectra were recorded in a raw extracts instead of pure compound and other co-eluting compounds could therefore interfere with the spectra, which led to some discrepancies in λ_{max} observations when comparing to previous literature

Figure 3.3 On opposite page: Extracted ion chromatograms (ESI–) of *S. cerevisiae* strains expressing type III PKSs. Relevant peaks representing compounds of interest are highlighted. (A) $m/z = 125.0233 \pm 0.01$, highlighted compound is triacetic acid lactone **(1)**; (B) $m/z = 191.0350 \pm 0.01$, highlighted compound is 5,7-dihydroxy-2-methylchromone **(2)**; (C) $m/z = 233.0455 \pm 0.01$, highlighted compound is 6-(2',4'-dihydroxy-6'-methylphenyl)-4-hydroxy-2-pyrone **(3)**; (D) $m/z = 275.0550 \pm 0.01$, highlighted compound is TW93a **(4)**; (E) $m/z = 231.0652 \pm 0.01$, highlighted compound is aloesone **(5)**; (F) $m/z = 317.0667 \pm 0.01$, highlighted compounds are SEK4 **(6)** and SEK4b **(7)** ; (G) $m/z = 299.0561 \pm 0.01$, highlighted compounds are dSEK4b **(8)** and dSEK4 **(9)**.


A
1
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]
B
3
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]
C
2
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]
D
4
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]
E
5
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]
F
6
7
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]
G
8
9
pBOSAL1-Gh2PS
pBOSAL1-AaPCS
pBOSAL1-DluHKS
pBOSAL1-AaPKS3
pBOSAL1-AaOKS
pBOSAL1
1.5
2.0
2.5
3.0
3.5
4.0
Time [min]

reporting the UV spectrum of the pure compound. An overview EICs with exact masses of each of the expected compounds can be found in figure 3.3 and UV and MS spectra of the tentatively identified compounds can be found in appendix (Figure S3.1).

Triacetic lactone (compound **1**) was expected to be produced by the Gh2PS PKS. The reported UV characteristic of a λ_{max} = 284nm [40] was also observed in our analysis along with an m/z = 125.0256 $[M\text{-}H]^-$. Also, a low RT of 1.4 min compared to other polyketides further supported the identification of compound **1**.

The expected pentaketide chromone (compound **2**) from AaPCS had an expected UV spectrum with λ_{max} = 228, 248, 256, 293 and 310nm [41]. The exact mass of m/z = 191.0348 $[M\text{-}H]^-$ (Calculated m/z = 191.0350) offered good support and the observed λ_{max} = 222, 248, 254sh, 294 also strongly supported the identification of the pentaketide chromone. The longer retention time of compound **3** (RT = 4.1 min) compared to the octaketides SEK4 (**6**, RT = 3.2 min) and SEK4b (**7**, RT = 3.3 min) was also in agreement with relative retention in reverse phase HPLC observed in a previous study by Abe *et al.* (2005) [28].

The hexaketide pyrone (compound **3**) had a λ_{max} = 302nm which was in good agreement with a previous study by Abe *et al.* (2005) [31]. Also, the detected exact mass of m/z = 233.0457 $[M\text{-}H]^-$ (Calculated m/z = 233.0455) supported the identification.

The strain expressing octaketide producing AaOKS produced higher levels of the heptaketide TW93a (compound **4**), than the strain expressing the heptaketide producing AaPKS3. TW93a has also been reported to be produced by AaOKS in easily detectable amounts in the study by Mizuuchi *et al.* (2009) [30]. TW93a had a reported λ_{max} = 216, 238 and 282 [42], which was in agreement with the observed λ_{max} = 216 and 278. Exact mass was determined to m/z = 275.0558 $[M\text{-}H]^-$ (calculated m/z = 275.0550) which supported the identification. An alternative compound identification of compound **4** could be the heptaketide 6-(2-(2,4- dihydroxy-6-methylphenyl)-2-oxoethyl)-4-hydroxy-2-pyrone (**10**), which has the same molecular formula. The heptaketide pyrone (**10**) almost co-elutes with TW93a (**4**) in the HPLC system where it is isolated, but in much lower concentration [30]. Therefore, it is most likely that the detected compound in our analysis is **4** as our UV-spectrum was more similar to the reported of **4** than what was reported for the heptaketide pyrone (**10**). Another heptaketide, aloesone (**5**), which

was reported to be produced by AaPKS3 had a reported λ_{max} = 243, 259 and 291 nm [43], and our observed UV-spectrum was in agreement with λ_{max} = 268 and 294 nm. The observed exact mass of m/z = 231.0660 $[M\text{-}H]^-$ (calculated m/z = 231.0652) also supported the identification of aloesone (**5**).

LC-MS and NMR based identification of compounds

The octaketides SEK4 and SEK4b both had a reported λ_{max} = 280 [28], which was also observed. The observed exact mass of m/z = 317.0663 (**6**) and m/z = 317.0668 (**7**) $[M\text{-}H]^-$ (calculated m/z = 317.0667 for both) supported the identification. The sequence of elution in the used reverse phase HPLC system for hexaketide, heptaketide and octaketide derived compounds are in agreement with the study by Mizuuchi *et al.* (2009) [30]. The dehydrated octaketides dSEK4b and dSEK4 (**8** and **9** respectively) were also detected in our chemical analysis, based on their exact mass and elution sequence when comparing with previous studies [18]. The observed exact mass of m/z = 299.0559 (**8**) and m/z = 299.0563 (**9**) (Calculated m/z = 299.0561 for both) supported their identification. No UV characteristics has been published for **8** and **9**, but there is a slight shift of λ_{max} = 280 nm for the non dehydrated octaketides SEK4 (**6**) and SEK4b (**7**) to a λ_{max} = 292 nm for the dehydrated octaketides dSEK4b (**8**) and dSEK4 (**9**). This can be explained with the higher number of conjugated double bonds in the dehydrated octaketides. An overview of the metabolites tentatively identified by HPLC-UV-MS and their characteristics in our chemical analysis system can be found in table 3.3. The octaketides **6**, **7**, **8** and **9** were purified by preparative chromatography and NMR analyses were done to verify the putative identifications of the octaketides based on HPLC-UV-MS. The observed NMR data was in agreement with previous studies [15, 37, 38]. NMR, MS and UV spectra from our analyses of the octaketides can be found in appendix (Figure S3.2 to S3.5).

A clear tendency in our efforts to express type III PKSs in *S. cerevisiae* was that the PKSs along with the longest possible polyketide chain also produced the shorter chains though in much lower titers than the longest polyketide chain. The titers were also much lower in comparison with PKS producing the polyketide as its longest polyketide chain. This effect has also been reported in previous studies of type III PKSs characterized under *in vitro* conditions [30]. An exception to this rule was DluHKS which was not found to produce the pentaketide **3**, but still was found to produce the triketide **1**. Also, the highest production of the heptaketide

Table 3.3 Above: The observed characteristics of detected metabolites. Below: Detection metabolites in strains.

Met #	RT	Molform	Calculated m/z [M-H]$^-$	Observed m/z [M-H]$^-$	Reported λ_{max} [nm]	Observed λ_{max} [nm]
1	1.4	$C_6H_6O_3$	125.0233	125.0256	284	288
3	4.1	$C_{10}H_8O_4$	191.0350	191.0353	228, 248, 256, 293, 310	222, 248, 254sh, 294
2	2.9	$C_{12}H_{10}O_5$	233.0455	233.0457	304	302
4	3.0	$C_{14}H_{12}O_6$	275.0550	275.0558	216, 238, 282	216, 278
5	3.7	$C_{13}H_{12}O_4$	231.0652	231.0666	243, 259, 291	268, 294
6	3.2	$C_{16}H_{14}O_7$	317.0667	317.0663	280	280
7	3.3	$C_{16}H_{14}O_7$	317.0667	317.0668	280	282
8	3.7	$C_{16}H_{12}O_6$	299.0561	299.0559	-	292
9	3.8	$C_{16}H_{12}O_6$	299.0561	299.0563	-	292

	Gh2PS	AaPCS	DluHKS	AaPKS3	AaOKS
1	+	+	+	+	+
3	nd	+	nd	+	+
2	nd	nd	+	+	+
4	nd	nd	nd	+	+
5	nd	nd	nd	+	+
6	nd	nd	nd	nd	+
7	nd	nd	nd	nd	+
8	nd	nd	nd	nd	+
9	nd	nd	nd	nd	+

4 was obtained by the octaketide producing AaOKS. This could be due to the aloesone producing AaPKS3 had a preference to produce the heptaketide aloesone **5**, instead of the heptaketide pyrone **4**. This is also in agreement with observations in the study by Mizuuchi *et al.* (2009) [30].

3.4.2 Co-expression of type III PKSs and type II cyclases

To test if it was possible to direct the folding of the produced polyketide backbone formed by the type III PKSs two bacterial polyketide cyclases were expressed by chromosomal integration in combination with the AaOKS type III PKS. The strain expressing the cyclases were of mating type α to enable crossing with strains expressing the octaketide synthase AaOKS as this were of mating type a. The chosen bacterial cyclases were PdmD from *Actinomadura hibisca* involved in pradimicin (Figure 2.3) biosynthesis [33] expected to promote the C9-C14 first ring folding pattern[44] and gra-orf4 from *Streptomyces vietnamensis* involved in granaticin (Figure 2.3) [32] expected to promote the C7-C12 first ring folding pattern. Only the octaketide producing type III PKS was chosen for co-expression with the cyclases as we assumed that this was the only PKS to produce a polyketide of a length, which would be foldable by both gra-orf4 and PdmD. The diploid strains expressing AaOKS and cyclases were created along with diploid control

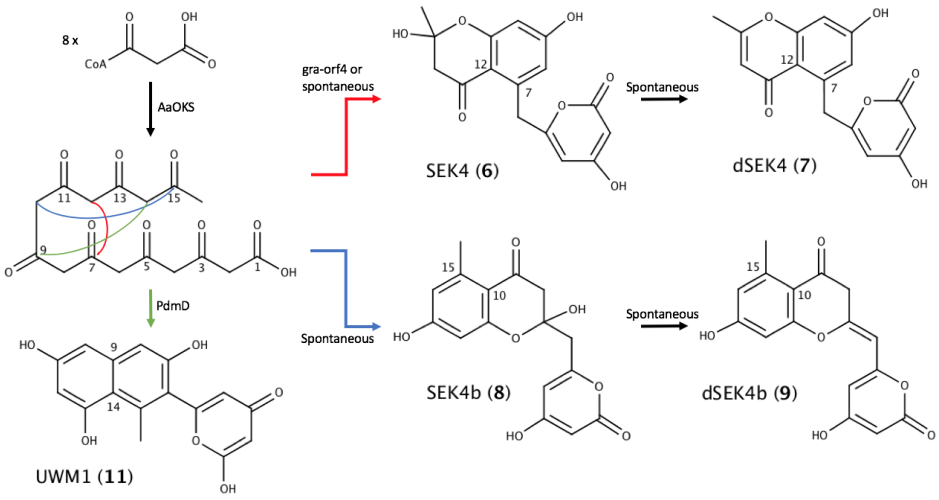


Figure 3.4 Compounds expected to be produced by co-expression of AaOKS and cyclases. The colored curved arrows indicate the first C-C bond formation which is induced spontaneously or by the cyclases gra-orf4 or PdmD.

strains expressing only the cyclases. These strains were found not to have altered metabolism compared to the parent strain (CEN.PK111-61A) not expressing the cyclases (data not shown).

The co-expression of AaOKS and Gra-orf4 was expected to increase the production of SEK4 (**6**) and dSEK4 (**9**) when comparing to the production of SEK4b (**7**) and dSEK4b (**8**). This was due to SEK4 (**6**) and dSEK4 (**9**) having the C7-C12 folding pattern promoted by the cyclase Gra-orf4, while SEK4b (**7**) and dSEK4b (**8**) has the C10-C15 spontaneous folding pattern. For the co-expression of the AaOKS and PdmD the expected compound to be promoted was the previously characterized UWM1 (compound **11**), which was observed when co-expressing the octaketide producing actinorhodin minimal PKS in combination with the tetracenomycin cyclase TcmN, which promotes the C9-C14 folding pattern [45]. An overview of the expected compounds, which would be produced by co-expression of the AaOKS and the cyclases, can be found in figure 3.4.

To evaluate the effects of the cyclases the area under peak for the different octaketides was measured for each strain grown in triplicate. The C7-C12 folded products were SEK4 and dSEK4 (compounds **6** and **9**), the C10-C15 folded products were SEK4b and dSEK4b (compound **7** and **8**), and the C9-C14 folded product was UWM1 (compound **11**). The percentage of the total production of oc-

Table 3.4 Production of octaketides by diploid strains grown in triplicate. The values are average percentages of the total area under peak for each of the folding patterns. The C7-C12% are representing SEK4 and dSEK4, C9-C14% are UVM1 and C10-C15 are representing SEK4b and dSEK4b. Standard deviations for triplicates are given in parentheses.

Enzymes	**OD_{600}**	**C7-C12 %**	**C9-C14 %**	**C10-C15 %**
AaOKS	9.87 (0.06)	47.77 (2.39)	1.09 (0.23)	51.14 (2.17)
AaOKS + Gra-orf4	8.49 (0.5)	54.02 (0.77)	0.19 (0.06)	45.8 (0.8)
AaOKS + PdmD	10.4 (0.96)	35.75 (0.36)	22.54 (2.71)	41.71 (2.35)

taketide for each of these folding patterns was calculated and these are shown in table 3.4. The identification of UWM1 (**11**) was only tentative as no previous studies have specified a UV spectrum for UWM1 (**11**). A strong support for the identification is the sequence of elution of SEK4, UWM1, and SEK4b, which is identical to the previous study by Seow *et al.* (1997) in which they also use reverse phase chromatography [14]. Also, an exact mass of $m/z = 299.0568$ (Calculated $m/z = 299.0561$) supported the tentative identification of the octaketide UWM1 (**11**). The elution sequence of the octaketides can be found in figure 3.5. The recorded MS and UV spectra for UWM1 (**11**) can be found in appendix (Figure S3.1).

The measured area under peak of octaketides (**6**, **7**, **8** and **9**) for each cultivation of the diploid strains expressing AaOKS with and without cyclases can be found in appendix (Table S3.2). It was found that the gra-orf4 cyclase was not able to fold all of the produced polyketide. Only about half of the produced octaketide was detected as the C7-C12 folded **6** and **9**. Nevertheless, the percentage of the total produced octaketide, which was folded in the C7-C12 folding pattern was significantly higher ($p = 0.049$) when co-expressing the C7-C12 cyclase. Therefore, this data shows that the cyclase gra-orf4 was able to direct the folding of the polyketide backbone. The co-expression of AaOKS and PdmD resulted in significantly higher levels of UWM1 (**11**) compared to the expression of AaOKS

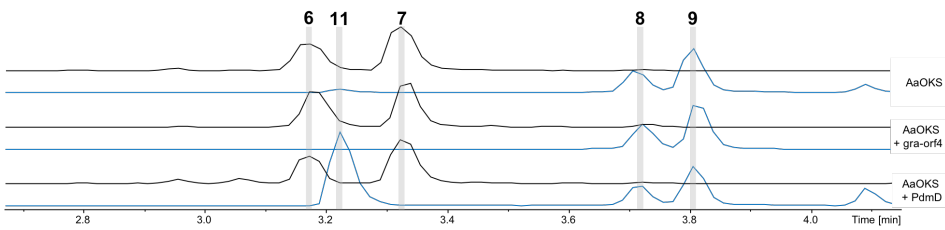


Figure 3.5 Detection of compounds to be quantified to evaluate folding patterns. Extracted ion chromatograms at $m/z = 317.0656 \pm 0.01$ (black) and $m/z = 299.0550 \pm 0.01$ (blue).

alone ($p = 0.006$). The change in the percentage was also much more prominent for the C9-C14 PdmD cyclase compared to the C7-C12 Gra-orf4 foldase. This shows that the nature of the foldases matters in terms of how efficient they are at accepting and converting the AaOKS octaketide into the final specific aromatic polyketide product.

When evaluating the efficiencies of the cyclases the natural substrate needs to be taken into account. In the natural pathway gra-orf4 folds an octaketide chain C_{16} [46], and issues with chain length of the AaOKS polyketide should not be present. The low efficiency originates from other factors such as non-functional or low expression, or high degradation rate of the gra-orf4 enzyme in *S. cerevisiae*. PdmD folds a dodecaketide (C_{24}) chain in the biosynthesis of pradicmicin [33], and it is therefore surprising that the PdmD cyclase is that efficient at folding the octaketide from AaOKS. The formation of the C6-C15 second ring is hypothesized to be spontaneous as this ring is not present in pradimicin, and thus not likely to be formed by PdmD.

3.5 Conclusion and perspectives

In this report we have shown that it is possible to express functional type III PKSs in *S. cerevisiae* that can produce an array of different polyketide backbones. Also, we have shown that it is possible to direct the first ring folding pattern of the polyketide backbone by co-expression of a type II cyclase with the polyketide directing the folding towards either the C7-C12 or C9-C14 folding pattern. Further direction of the folding pattern of the polyketide backbone by having second and/or third ring folding cyclases in the platform, would enable production of a vast array of compounds. This has been successfully done in *Aspergillus nidulans* previously for the production of the industrial colorant carmine [18]. The work presented here functions as a proof-of-concept setting the scene for using a wider range of type III PKSs producing longer polyketide products. In this way, the range of polyketides the platform would be able to produce would grow fast as longer chains can be folded in more complex ways. Also, it would be interesting to test cyclases from other type II PKS systems, as it may be possible to find cyclases which would be more efficient in producing the C7-C12 or C9-C14 folding pattern, or enzymes that catalyze other cyclisation patterns.

3.6 Acknowledgements

We would also like to thank the DTU Bioengineering Metabolomics platform for assistance in running the HPLC-UV-MS analysis, and the NMR center DTU at DTU Chemistry for assistance in running the NMR analyses. This work was funded by the Novo Nordisk foundation grant number NNF15OC0016626.

3.7 Author contributions

The research study was conceived by RF and HC. The molecular biology work was designed and conducted by HC and CS. The cultivations and chemical work was done by KK. Drafting of the manuscript was done by KK, and refined by RF and TL.

References

[1] Butler, MS. "Natural products to drugs: natural product-derived compounds in clinical trials". In: *Nat Prod Rep* 25.3 (2008), pp. 475–516.

[2] Dewick, PM. *Medicinal Natural Products - A Biosynthetic Approach, Third Edition*. 3rd. John Wiley & Sons Ltd, 2009.

[3] Weissman, KJ. "Uncovering the structures of modular polyketide synthases". In: *Nat Prod Rep* 32.3 (2015), pp. 436–453.

[4] Cox, RJ and Simpson, TJ. "Fungal Type I Polyketides". In: *Comprehensive Natural Products II - Chemistry and Biology*. Ed. by H Liu and L Mander. Vol. 1. Elsevier, 2010, pp. 347–383.

[5] Hertweck, C, Luzhetskyy, A, Rebets, Y, and Bechthold, A. "Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork". In: *Nat Prod Rep* 24.1 (2007), pp. 162–190.

[6] Abe, I and Morita, H. "Structure and function of the chalcone synthase superfamily of plant type III polyketide synthases". In: *Nat Prod Rep* 27.6 (2010), pp. 809–838.

[7] Hertweck, C. "The biosynthetic logic of polyketide diversity". In: *Angew Chem Int Ed* 48.26 (2009), pp. 4688–4716.

[8] Crawford, JM and Townsend, CA. "New insights into the formation of fungal aromatic polyketides". In: *Nat Rev Microbiol* 8.12 (2010), p. 879.

[9] Zhou, H, Li, Y, and Tang, Y. "Cyclization of aromatic polyketides from bacteria and fungi". In: *Nat Prod Rep* 27.6 (2010), pp. 839–868.

[10] Gagne, SJ, Stout, JM, Liu, E, Boubakir, Z, Clark, SM, and Page, JE. "Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides". In: *Proc Nat Acad Sci* 109.31 (2012), pp. 12811–12816.

[11] Xu, Y, Zhou, T, Zhou, Z, Su, S, Roberts, SA, Montfort, WR, Zeng, J, Chen, M, Zhang, W, Lin, M, Zhan, J, and Molnar, I. "Rational reprogramming of fungal polyketide first-ring cyclization". In: *Proc Nat Acad Sci* 110.14 (2013), pp. 5398–5403.

[12] Menzella, HG, Carney, JR, and Santi, DV. "Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases". In: *Cell Chem Biol* 14.2 (2007), pp. 143–151.

[13] Newman, AG, Vagstad, AL, Storm, PA, and Townsend, CA. "Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming". In: *J Am Chem Soc* 136.20 (2014), pp. 7348–7362.

[14] Seow, KT, Meurer, G, Gerlitz, M, Wendt-Pienkowski, E, Hutchinson, CR, and Davies, J. "A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: A means to access and use genes from uncultured microorganisms". In: *J Bacteriol* 179.23 (1997), pp. 7360–7368.

[15] Fu, H, Ebert-Khosla, S, Hopwood, DA, and Khosla, C. "Engineered biosynthesis of novel polyketides: dissection of the catalytic specificity of the act ketoreductase". In: *J Am Chem Soc* 116.10 (1994), pp. 4166–4170.

[16] Zhang, W, Li, Y, and Tang, Y. "Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*". In: *Proc Nat Acad Sci* 105.52 (2008), pp. 20683–20688.

[17] Andersen-Ranberg, J, Kongstad, KT, Nafisi, M, Staerk, D, Okkels, FT, Mortensen, UH, Lindberg Møller, B, Frandsen, RJN, and Kannangara, R. "Synthesis of C-Glucosylated Octaketide Anthraquinones in *Nicotiana benthamiana* by Using a Multispecies-Based Biosynthetic Pathway". In: *ChemBioChem* 18.19 (2017), pp. 1893–1897.

[18] Frandsen, RJ, Khorsand-Jamal, P, Kongstad, KT, Nafisi, M, Kannangara, RM, Staerk, D, Okkels, FT, Binderup, K, Madsen, B, Møller, BL, Thrane, U, and Mortensen, UH. "Heterologous production of the widely used natural food colorant carminic acid in *Aspergillus nidulans*". In: *Sci Rep* 8.12853 (2018), pp. 1–10.

[19] Walker, GM and Stewart, GG. "*Saccharomyces cerevisiae* in the production of fermented beverages". In: *Beverages* 2.30 (2016), pp. 1–12.

[20] Walsh, G. "Biopharmaceutical benchmarks 2014". In: *Nat Biotechnol* 32.10 (2014), pp. 992–1000.

[21] Entian, KD and Kötter, P. "25 Yeast Genetic Strain and Plasmid Collections". In: *Methods Microbiol* 36.06 (2007), pp. 629–666.

[22] Frandsen, RJN, Mortensen, UH, Coumou, HC, Kannangara, RM, Madsen, B, Nafisi, M, Andersen-Ranberg, J, Kongstad, KT, Okkels, FT, Khorsand-Jamal, P, and Stærk, D. "Use of heterologous expressed polyketide synthase and small molecule foldases to make aromatic and cyclic compounds". IPC No. C12P 19/60 A I. Patent No. WO2016198623. 2016.

[23] Jensen, NB, Strucko, T, Kildegaard, KR, David, F, Maury, J, Mortensen, UH, Forster, J, Nielsen, J, and Borodina, I. "EasyClone: Method for iterative chromosomal integration of multiple genes in *Saccharomyces cerevisiae*". In: *FEMS Yeast Res* 14.2 (2014), pp. 238–248.

[24] Sherman, F, Fink, GR, and Hicks, JB. *Laboratory Course Manual for Methods in Yeast Dynamics*. Cold Spring Harbor Laboratory, 1986.

[25] Woodcock, DM, Crowther, PJ, Doherty, J, Jefferson, S, DeCruz, E, Noyer-Weidner, M, Smith, SS, Michael, MZ, and Graham, MW. "Quantitative evaluation of *Escherichia coli* host strains for tolerance to cytosine methylation in plasmid and phage recombinants". In: *Nucleic Acids Res* 17.9 (1989), pp. 3469–3478.

[26] Nørholm, MHH. "A mutant *Pfu* DNA polymerase designed for advanced uracil-excision DNA engineering". In: *BMC Biotechnol* 10 (2010).

[27] Helariutta, Y, Elomaa, P, Kotilainen, M, Griesbach, RJ, Schröder, J, and Teeri, TH. "Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (Asteraceae)". In: *Plant Mol Biol* 28 (1995), pp. 47–60.

[28] Abe, I, Utsumi, Y, Oguro, S, Morita, H, Sano, Y, and Noguchi, H. "A plant type III polyketide synthase that produces pentaketide chromone". In: *J Am Chem Soc* 127.5 (2005), pp. 1362–1363.

[29] Jindaprasert, A, Springob, K, Schmidt, J, De-Eknamkul, W, and Kutchan, TM. "Pyrone polyketides synthesized by a type III polyketide synthase from Drosophyllum lusitanicum". In: *Phytochem* 69.18 (2008), pp. 3043–3053.

[30] Mizuuchi, Y, Shi, SP, Wanibuchi, K, Kojima, A, Morita, H, Noguchi, H, and Abe, I. "Novel type III polyketide synthases from *Aloe arborescens*". In: *FEBS J* 276.8 (2009), pp. 2391–2401.

[31] Abe, I, Oguro, S, Utsumi, Y, Sano, Y, and Noguchi, H. "Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase". In: *J Am Chem Soc* 127.36 (2005), pp. 12709–12716.

[32] Deng, MR, Guo, J, Li, X, Zhu, CH, Zhu, HH, Deng, MR, Guo, ÁJ, Zhu, CH, Zhu, HH, and Li, X. "Granaticins and their biosynthetic gene cluster from *Streptomyces vietnamensis*: evidence of horizontal gene transfer". In: *Antonie van Leeuwenhoek* 100 (2011), pp. 607–617.

[33] Kim, BC, Lee, JM, Ahn, JS, and Kim, BS. "Cloning, sequencing, and characterization of the pradimicin biosynthetic gene cluster of Actinomadura hibisca P157-2". In: *J Microbiol Biotechnol* 17.5 (2007), pp. 830–839.

[34] Gietz, RD and Woods, RA. "Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method". In: *Method Enzymol* 350 (2002), pp. 87–96.

[35] Verduyn, C, Postma, E, Scheffers, WA, and van Dijken, JP. "Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation". In: *Yeast* 8.7 (1992), pp. 501–517.

[36] Pronk, JT. "Auxotrophic yeast strains in fundamental and applied research". In: *Appl Environ Microbiol* 68.5 (2002), pp. 2095–2100.

[37] Fu, H, Hopwood, DA, and Khosla, C. "Engineered biosynthesis of novel polyketides: evidence for temporal, but not regiospecific, control of cyclization of an aromatic polyketide precursor". In: *Cell Chem Biol* 1.4 (1994), pp. 205–210.

[38] Xiang, L, Kalaitzis, JA, and Moore, BS. "EncM, a versatile enterocin biosynthetic enzyme involved in Favorskii oxidative rearrangement, aldol condensation, and heterocycle-forming reactions". In: *Proc Nat Acad Sci* 101.44 (2004), pp. 15609–15614.

[39] Cardenas, J and Da Silva, NA. "Metabolic engineering of *Saccharomyces cerevisiae* for the production of triacetic acid lactone". In: *Met Eng* 25 (2014), pp. 194–203.

[40] Yamada, K. "Infrared and ultraviolet spectra of α-and γ-pyrones". In: *Bull Chem Soc Jpn* 35.8 (1962), pp. 1323–1329.

[41] Brown, RT, Blackstock, WP, and Chapple, CL. "Isolation of 5, 7-dihydroxy-2-methylchromone and its 7-O-glycosides from Adina rubescens". In: *J Chem Soc* 18 (1975), pp. 1776–1778.

[42] Yu, TW, Shen, Y, McDaniel, R, Floss, HG, Khosla, C, Hopwood, DA, and Moore, BS. "Engineered biosynthesis of novel polyketides from Streptomyces spore pigment polyketide synthases". In: *J Am Chem Soc* 120.31 (1998), pp. 7749–7759.

[43] Abe, I, Utsumi, Y, Oguro, S, and Noguchi, H. "The first plant type III polyketide synthase that catalyzes formation of aromatic heptaketide". In: *FEBS lett* 562.1-3 (2004), pp. 171–176.

[44] Zhan, J, Watanabe, K, and Tang, Y. "Synergistic actions of a monooxygenase and cyclases in aromatic polyketide biosynthesis". In: *ChemBioChem* 9.11 (2008), pp. 1710–1715.

[45] Shen, B, Summers, RG, Wendt-Pienkowski, E, and Hutchinson, CR. "The *Streptomyces glaucescens tcmKL* polyketide synthase and *tcmN* polyketide cyclase genes govern the size and shape of aromatic polyketides". In: *J Am Chem Soc* 117.26 (1995), pp. 6811–6821.

[46] Ichinose, K, Bedford, DJ, Tornus, D, Bechthold, A, Bibb, MJ, Revill, WP, Floss, HG, and Hopwood, DA. "The granaticin biosynthetic gene cluster of Streptomyces violaceoruber Tü22: sequence analysis and expression in a heterologous host". In: *Cell Chem Biol* 5.11 (1998), pp. 647–659.

4 Manuscript II – Production of flavokermesic acid in yeast by enzyme fusion

Kromphardt KJK[a], Skovbjerg CAS[a], Vestergaard AM[a], Rasmussen R[a], Hagerup T[a], Larsen TO[a] & Frandsen RJN[a]
[a] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark

This manuscript is planned for submission to Scientific Reports.

4.1 Abstract

Carmine is an industrial colorant which has long been recognized for its industrial importance as it has many favorable traits for industrial use. The current production of carmine is vulnerable to uncontrollable factors and thus an artificial biosynthetic pathway for production of the carminic acid precursor flavokermesic acid has been successfully expressed in *Nicotiana benthamiana* and *Aspergillus nidulans*. However, the production of shunt products in these organisms was found to be high and an alternative method is therefore needed. We aimed firstly to express the pathway in *Saccharomyces cerevisiae* to investigate the performance of the pathway in *S. cerevisiae* and secondly to investigate if translational fusion of the biosynthetic enzymes would improve output of the pathway and reduce shunt products. We found that *S. cerevisiae* was able to produce 52 mg flavokermesic acid per liter culture while the fusion of the biosynthetic enzymes was found to have a detrimental effect on the performance of the artificial biosynthetic pathway. Proteomics analysis of the cultivated strains revealed that fused enzymes were in much lower relative abundance compared to non fused enzymes. This lead us to hypothesize that fusion of the biosynthetic enzymes interfered with their ability to form dimers. The creation of a *S. cerevisiae* cell factory producing flavokermesic acid serves as an important step towards creating a stable fermentation based production of carminic acid.

Keywords: Yeast, Metabolic engineering, Polyketide synthases

4.2 Introduction

The industrially important natural colorant carmine is derived from carminic acid which is extracted from the scale insect *Dactylopius coccus* along with related compounds such as flavokermesic acid (FK) and kermesic acid (KA) [1]. Carmine has reached this importance because of its solubility, coloring potential and stability, and humans have a long history of using the compound as dye for textiles, paints, food and cosmetics [2]. The current production method for carmine is vulnerable to a series of uncontrollable factors that typically plague agricultural products, such as the weather and prevalence of plant or insect pathogens, which result in large yearly fluctuations in the available carmine levels, leading to a very unstable market and pricing. An increasing demand for natural colors from consumers has further increased pricing of carmine. To circumvent this, a fermentable cell factory would stabilize the supply and increase the production levels to meet demands. However, the natural biosynthetic pathway and the involved enzymes in *D. coccus* is currently unknown [3], and it has been necessary to build an artificial biosynthetic pathway [4, 5]. The three initial steps of the artificial biosynthetic pathway in question are shown in figure 4.1 and has been shown to lead to the production of

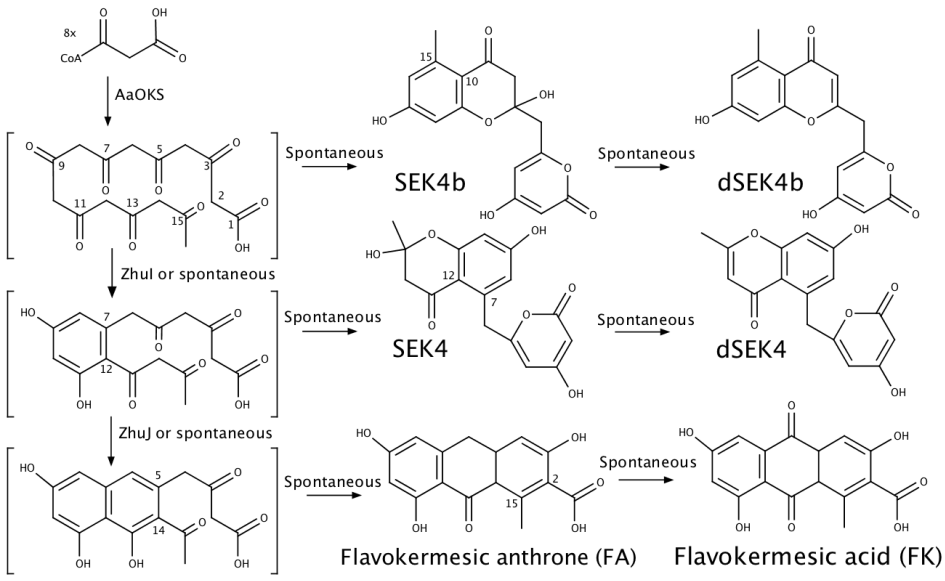


Figure 4.1 The artificial biosynthetic pathway for the production of flavokermesic acid, including possible spontaneous reactions and products. The carbons in the structures are numbered based on their position in the octaketide chain.

flavokermesic acid in *Nicotiana benthamiana*[4] and when expressed in *Aspergillus nidulans* also kermesic acid [5].

The artificial biosynthetic pathway for production of flavokermesic acid was created by combining enzymes from a plant type III PKS and a bacterial type II PKS systems. The pathway is initiated by the *Aloe arborescens* octaketide synthase (OKS), which produces an non-reduced polyketide consisting of 16 carbons [6]. This unstable octaketide can spontaneously fold to form flavokermesic acid, but with very low yields due to multiple competing spontaneous reactions creating shunt products. The desired reactions can, however, be promoted by enzymes catalyzed formation of the C7-C12 bond in the molecule, using the cyclase ZhuI, and formation of the second C5-C14 bond can likewise be promoted using the cyclase ZhuJ[1]. The two cyclases originate from *Streptomyces* sp. R1128, where they participate in biosynthesis of anthraquinones from a octaketide chain [7]. Formation of the third C2-C15 bond that forms flavokermesic anthrone, and the subsequent oxidation of the central ring to form the quinone of flavokermesic acid, are both hypothesized to be spontaneous reactions in *N. benthamiana* and *A. nidulans* [4, 5].

The production levels and product spectra reported in the two proof-of-concept studies for the pathway [4, 5], however, showed that the pathway in its current form is inefficient. This is primarily due to a high level of shunt reactions leading to the formation of e.g. SEK4 and SEK4B instead of the desired flavokermesic acid and kermesic acid. A situation that makes it futile to attempt to optimize production to industrially relevant levels simply by increasing the precursor supply in the cell or the enzyme levels as this would not mend the problem of shunt product formation.

In the current study we test different strategies to optimize the specificity of the artificial biosynthetic pathway by first expressing it in *Saccharomyces cerevisiae*, which is widely recognized and utilized in academia and industry for the development and production of small molecules and proteins. *S. cerevisiae* features a less complex metabolism and reactome compared to the previously tested hosts, which reduces the chances of shunt product formation via competing endogenous pathways. Secondly, we investigated if fusion of the biosynthetic enzymes could increase the efficiency and specificity of the pathway and allow for elevated levels

[1]The polyketide fold catalyzing enzymes ZhuI and ZhuJ have vastly different domain structure and the names cyclase and aromatase are used for the respectively, but for ease of reading they are both called cyclase in this manuscript.

of flavokermesic acid.

4.3 Materials and methods

4.3.1 Synthetic genes, vectors and strains, and media

Biosynthetic genes *OKS* (Q3L7F5, [6]), *ZhuI* (AAG30197, [7]), and *ZhuJ* (AAG30196, [7] were codon optimized for *S. cerevisiae* and *de novo* synthesized by Genscript (http://www.genscript.com). The three enzymes were translationally fused in different configurations via two different amino acid linkers, here called L1 (Sequence: MGSGGGGS), designed by Albertsen *et al.* (2011) [8], and amino acid linker L2 (Sequence: MGGSGSAG) which was designed for this study based on the design rules for designing L1, by utilizing small polar amino acids, such as serine and glycine, which ensures that the linker is flexible and water soluble [8]. The yeast integrative expression vectors used in this study were derived from the system described by Mikkelsen *et al.* (2012) [9], and a complete list of the used and constructed vectors can be found in appendix (Table S4.2). *Escherichia coli* strain DH5α was used to assemble and propagate all plasmids. An overview of all used and created *S. cerevisiae* strains from this study can be found in appendix (Table S4.3). Lysogeny broth (LB) medium supplemented with 100 µg/L ampicillin was used for all *E. coli* vector propagations. Synthetic complete (SC) medium lacking uracil [10] was used for selecting yeast transformants based on acquisition of the URA3 marker in the introduced expression cassette. Subsequent propagation of transformants was performed on yeast peptone dextrose (YPD) medium consisting of 10 g/L Bacto yeast extract (BD Bionutrients, NJ, USA), 20 g/L Bacto Peptone (BD Bionutrients, NJ, USA) and 20 g/L glucose. To allow for consecutive transformation steps in the yeast strains, the introduced URA3 marker was removed via direct repeat recombination and counter selection for the URA3 marker by growing transformants on SC supplemented with 740 mg/L 5-fluoroorotic acid (USBiological, MA, USA).

4.3.2 Vector and strain construction

The coding sequence of the three genes were PCR amplified from the plasmids delivered from GenScript using the combinations of primers listed in appendix (Table S4.1). As up to three enzymes would need to be fused with the linkers in between, three positions (beginning (B), middle (M) and end (E)) was designated. Directional assembly was possible as universal primer overhangs containing in the

linkers (L1 or L2) were designed. The yeast promoters P_{TEF1} and P_{PGK1} were PCR amplified from pSP2 [11] using primers also with universal overhangs. These universal overhangs allowed for directional seamless DNA assembly. The yeast expression vectors were assembled via the Uracil-Specific Excision Reagent (USER) fusion cloning method as described in Mikkelsen *et al.* (2012) [9]. The resulting plasmids were verified by restriction analysis and sanger sequencing (GATC biotech, Konstanz, Germany). The individual expression cassettes were excised from the vectors via restriction enzyme digestion and gel-purified prior to introduction into yeast, as described in Mikkelsen *et al.* (2012) [9]. Transformation of *S. cerevisiae* was performed using the LiAc/SS-DNA/PEG method [12]. Transformants with the desired genotype were verified via diagnostic colony PCR using primers specific for the integrated genes and the backbone (Colony PCR/Sequencing primers in table S4.1).

4.3.3 Cultivation and chemical extraction

S. cerevisiae strains were cultured in triplicates for 5 days at 30°C in 500 mL Erlenmeyer flasks with bafles containing 50mL YPD at 200 rpm. All cultures were innoculated with an overnight YPD preculture to an OD_{600} of 0.1. OD_{600} measurements were performed using a Shimadzu UV-1800 spectrophotometer (Japan). Following completion of the cultivation the end-point OD_{600} was measured and 1 mL of the culture was removed for proteomics analysis, and the remaining was used for the metabolite analysis. The cells in the proteomics sample were collected by centrifugation at 8000 xg and washed once with milliQ water and stored at -18°C until all samples were ready for proteomics analysis.

Chemical extractions were done by separating the biomass from the culture liquid via centrifugation (8000 xg for 5 minutes) in a 50mL falcon tube (Subsequently named tube A). The culture liquid was decanted into another 50mL falcon tube (Subsequently named tube B). To lyse the yeast cells 25mL of extraction liquid (ethylacetate with 1% v/v formic acid) was added to the pellet in tube A and subjected to ultrasound for 1 hour. Afterwards, 25mL of the culture liquid from tube B was decanted into the extraction liquid and yeast cell mixture. After vigorous shaking the culture liquid was removed with a glass Pasteur pipette from tube A. Tube A was centrifuged (8000 xg for 5 minutes) and the extraction liquid was decanted into the remaining culture liquid tube B. After vigorous shaking with the remaining culture liquid the culture liquid was removed with a

glass Pasteur pipette. The extraction liquid was evaporated to dryness under a gentle stream of N_2. The residue was redissolved in 300µL HPLC grade methanol and exposed to ultrasound for 10 seconds to dissolve residuals and transferred to a 2 mL eppendorph tube. The eppendorph tube was centrifuged (18000 xg for 5 minutes) to remove undissolved residuals and 150µL of the liquid was transferred to a HPLC-vial.

4.3.4 HPLC-MS quantification

Samples were separated on an Agilent Infinity 1290 UHPLC (Agilent Technologies, CA, USA) fitted with a Kinetex Biphenyl (100 x 2.1 mm, 2.6 µm) column kept at 40°C. The gradient elution was performed with a solvent A consisting of H_2O with 50mM formic acid and a solvent B consisting of acetonitrile with 50mM formic acid. The gradient profile was as follows: 15% B and kept there for 0.5 min, 15% B to 25% B in 1.5 min, 25% B to 25.56% B in 3.5 min, 25.56% B to 100% B in 0.5 min kept at 100% for 1.5 min and returned to 15% B in 0.1 min and equilibrated for 1.9 min with a flow of 0.4 mL/min. The sample injection volume was 1µL. The MS quantification was performed on an Agilent 6490 Triple Quadropole MS fitted with an electrospray ionisation (ESI) ion source. The source settings were as follows: gas temp 200°C, gas flow 16 l/min, nebulizer 25 psi, sheath gas temp 300°C sheath gas flow 12 l/min, capillary voltage 4000V and nozzle voltage 500V. iFunnel parameters were at High pressure RF of 150V and low pressure RF of 60V. To quantify the compounds of interest the parent masses were the $[M\text{-}H]^-$ for the individual compounds, with collision energies of 10, 20 and 40V, and the quantified product ions as follows: SEK4 (m/z 233, RT = 2.05 min), SEK4b (m/z 191, RT = 2.17 min), dSEK4 (m/z 213, RT = 2.78 min), dSEK4b (m/z 83, RT = 2.62 min) and flavokermesic acid (m/z 269, RT = 3.90 min). Standards of each of the quantified compounds were used for a calibration curve (0.1-100 µg/L) with only the linear part being used for quantification. Cultivation samples were diluted (5, 10 or 20 times with methanol) if needed. HPLC-MS data analysis was performed in Agilent QQQ Quantitative analysis (Agilent Technologies, CA, USA).

4.3.5 Proteomics analysis

Lysate preparation and digestion was done according to Kulak *et al.* (2014) [13]. Briefly, cells were lysed using 50µl of lysis buffer (consisting of 6 M Guanidinium

Hydrochloride, 10 mM TCEP, 40 mM CAA, 100 mM Tris pH8.5). Samples were boiled at 95°C for 5 minutes, after which they were sonicated on high for 3 times 10 seconds in a Bioruptor sonication water bath (Diagenode) at 4°C. After determining protein concentration with Bradford (Sigma), 10ug was taken forward for digestion. Samples were diluted 1:3 with 10% acetonitrile, 25 mM Tris pH 8.5, LysC (MS grade, Wako) was added in a 1:50 (enzyme to protein) ratio, and samples were incubated at 37° for 4 hours. Samples were further diluted to 1:10 with 10% acetonitrile, 25 mM Tris pH 8.5, trypsin (MS grade, Promega) was added in a 1:100 (enzyme to protein) ratio and samples were incubated overnight at 37°. Enzyme activity was quenched by adding 2% trifluoroacetic acid (TFA) to a final concentration of 1%. Prior to TMT labeling, the peptides were desalted on in-house packed C18 Stagetips [14]. For each sample, 2 discs of C18 material (3M Empore) were packed in a 200µl tip, and the C18 material activated with 40µl of 100% methanol (HPLC grade, Sigma), then 40µl of 80% acetonitrile, 0.1% formic acid. The tips were subsequently equilibrated 2x with 40µl of 1%TFA, 3% acetonitrile, after which 10ug of sample was loaded using centrifugation at 4,000x rpm. After washing the tips twice with 100µl of 0.1% formic acid, the peptides were eluted into clean 500µl Eppendorf tubes using 40% acetonitrile, 0.1% formic acid. The eluted peptides were concentrated in an Eppendorf Speedvac, and re-constituted in 50mM HEPES (pH8.5) for TMT labeling. Labeling was done according to manufacturer's instructions, and subsequently, labeled peptides were mixed 1:1:1:1:1:1:1:1:1:1:1 (11-plex), acidified to 1% TFA and acetonitrile concentration brought down to <5% using 2% TFA. Prior to mass spectrometry analysis, the peptides were fractionated using Strong Cation Exchange (SCX) in StageTip format. For each sample, 3 discs of SCX material (3M Empore) were packed in a 200µl tip, and the SCX material activated with 80µl of 100% acetonitrile (HPLC grade, Sigma). The tips were equilibrated with 80µl of 0.2% TFA, after which the samples were loaded using centrifugation at 4,000x rpm. After washing the tips twice with 100µl of 0.2% TFA, 2 initial fractions were eluted into clean 500µl Eppendorf tubes using 75 and 200mM ammonium acetate in 20% acetonitrile, 0.5% formic acid respectively. The final fraction was eluted using 5% ammonium hydroxide, 80% acetonitrile. The eluted fractions were frozen on dry ice and concentrated in an Eppendorf Speedvac, and re-constituted in 1% TFA, 2% acetonitrile containing iRT peptides for Mass Spectrometry (MS) analysis.

For each sample, peptides were loaded onto a 2 cm C18 trap column (Ther-

moFisher 164705), connected in-line to a 50 cm C18 reverse-phase analytical column (Thermo EasySpray ES803) using 100% Buffer A (0.1% Formic acid in water) at 750 bar, using the Thermo EasyLC 1000 HPLC system, and the column oven operating at 45°C. Peptides were eluted over a 140 minute gradient ranging from 6 to 60% of 80% acetonitrile, 0.1% formic acid at 250 nl/min, and the Q-Exactive instrument (Thermo Fisher Scientific) was run in a DD-MS2 top10 method. Full MS spectra were collected at a resolution of 70000, with an AGC target of 3 x 10^6 or maximum injection time of 20 ms and a scan range of 300–1750 m/z. The MS2 spectra were obtained at a resolution of 35000 and an AGC target value of 1 x 10^6 or maximum injection time of 120 ms, a normalised collision energy of 28 and an intensity threshold of 8.3 x 10^3. Dynamic exclusion was set to 60 s, and ions with a charge state < 2 or unknown were excluded. MS performance was verified for consistency by running complex cell lysate quality control standards, and chromatography was monitored to check for reproducibility.

The raw files were analysed using Proteome Discoverer 2.2. TMT reporter ion quantitation was enabled in the processing and consensus steps, and spectra were matched against the CEN.PK113-7D (proteome ID UP000013192) database obtained from Uniprot, with sequences of the heterologously expressed enzymes added (See appendix page 132 for the added sequences). Dynamic modifications were set as Oxidation (M), Deamidation (N,Q) and Acetyl on protein N-termini. Cysteine carbamidomethyl was set as a static modification. All results were filtered to a 1% FDR, and protein quantitation done using the built-in Minora Feature Detector, combined with the built-in ANOVA feature for statistical significance testing.

4.3.6 Analysis of protein structures

To investigate the protein structures the crystal structure for OKS and ZhuI was downloaded and visualized in PyMOL [15]. As no crystal structure for ZhuJ had been published, a homology model was build by utilizing HHPRED followed by homology modelling with the best sequence homologs. The homolog search and modelling was done with the Bioinformatics toolkit (Available at https://toolkit.tuebingen.mpg.de) [16].

4.4 Results and discussion

4.4.1 Expression of artificial flavokermesic acid producing biosynthetic pathway in *S. cerevisiae*

To investigate the functionality of the artificial flavokermesic acid biosynthetic pathway in *S. cerevisiae*, we first constructed a series of yeast strains containing the three individual genes encoding the biosynthetic enzymes in all possible combinations. These were OKS, OKS + ZhuI, OKS + ZhuJ and OKS + ZhuI + ZhuJ, along with control strains only expressing the cyclases ZhuI and ZhuJ. The expression of the artificial pathway producing flavokermesic acid was found to produce flavokermesic acid along with the shunt products SEK4, SEK4b, dSEK4 and dSEK4b (Figure 4.2). The control strains which were not expressing OKS did not produce any detectable amounts of the metabolites quantified in the strains expressing OKS. Mutactin was detected in *A. nidulans* when expressing the artificial pathway [5], but mutactin was not detected when expressing the pathway in *S. cerevisiae*.

The production of flavokermesic acid was found to be highest by the strain expressing all three biosynthetic enzymes with a production of 52.26 mg flavokermesic acid per liter culture (Table 4.1). While this was a high yield it was also positive to conclude that as good as 100% of the detected metabolites was flavokermesic acid meaning that the system was very efficient at producing flavokermesic

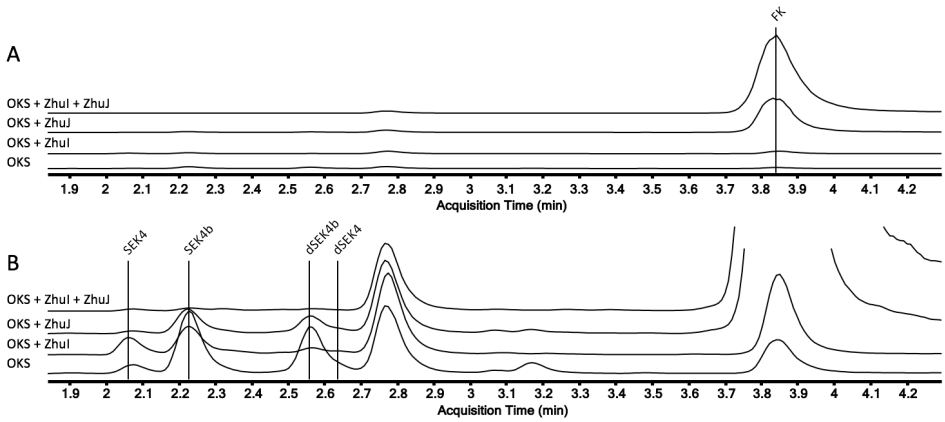


Figure 4.2 Total ion chromatogram (TIC) traces of strains expressing the artificial biosynthetic pathway. (A) An overview of the TIC chromatogram highlighting production of flavokermesic acid. (B) Same chromatogram as A but scaled differently to view production of shunt products SEK4, SEK4b, dSEK4b, dSEK4.

Table 4.1 Metabolic performance of the strains expressing the artificial pathway producing flavokermesic acid. Total is the summed amount of metabolites SEK4, SEK4b, dSEK4b, dSEK4, and FK in mg per litre culture. FK is the specific production of flavokermesic acid in mg per litre culture. C7-C12 % is the percentage of the total metabolites, which had the C7-C12 folding pattern. FK% is the percentage of total metabolites being FK.

Enzymes	Total	FK	C7-C12 %	FK %
OKS	0,44 (0,27)	0,24 (0,06)	36,45 (10,6)	28,57 (13,08)
OKS + ZhuI	1,21 (0,3)	0,57 (0,13)	72,82 (2,8)	47,42 (5,99)
OKS + ZhuJ	12,36 (1,82)	12,09 (1,67)	97,96 (0,85)	97,84 (0,9)
OKS + ZhuI + ZhuJ	52,26 (4,16)	52,23 (4,17)	99,95 (0,04)	99,95 (0,04)

acid with very few shunt products. Previous studies did not report quantification of metabolites, which made direct comparison of cell factories troublesome, but with very limited production of shunt products compared to flavokermesic acid the artificial biosynthetic pathway appeared to be performing substantially better in *S. cerevisiae* than in *N. benthamiana* and *A. nidulans* [4, 5]. In previous studies the level of flavokermesic acid was comparable or lower than the shunt products [4, 5], whereas we found that the level of flavokermesic acid was more than 2000-fold that of shunt products in the strain expressing OKS, ZhuI and ZhuJ.

With this positive first expression in *S. cerevisiae* of the artificial pathway for production of flavokermesic acid we explored methods for optimizing the pathway in *S. cerevisiae* to obtain even higher production of flavokermesic acid. Fusion of biosynthetic pathway enzymes has previously been shown to improve pathway turnover and specificity because the intermediates released from one enzyme are more likely to diffuse to the next enzyme in the pathway if these are fused and thus in close proximity [17]. We chose to utilize flexible, water soluble and relatively long linkers (8 amino acids) [8]. As we did not know how the enzymes would react to having the linkers and other enzymes attached, we wanted to allow the individual enzymes to move freely in reference to each other. Furthermore, as we did not know what combination or sequence of the enzymes would be best for production of flavokermesic acid when fused, we would have to investigate all possible combinations of fusions of two and three enzymes.

4.4.2 Effect of different translation fusion of OKS, ZhuI and ZhuJ on flavokermesic acid production

The construction of translational fusion required that the OKS would be fused to the other enzymes, which might impact its performance. We therefore tested the effects of appending the L1 and L2 linkers to the OKSs N- and C-terminal ends, to learn the if linkers would effect the function of the OKS. Not only would this show if

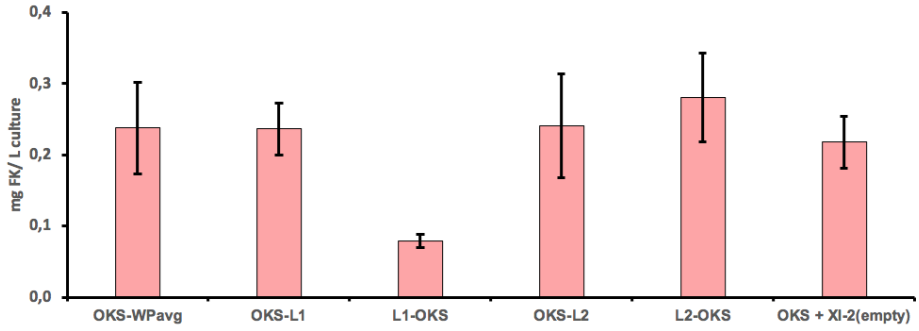


Figure 4.3 Production of Flavokermesic acid by strains expressing the OKS only.

the linkers could interfere with the active site, but as the OKS functions as a dimer it would also give an indication if the linker interfered with the dimeric space. The level of flavokermesic acid was comparable for all strains except for the construct expressing the L1 linker linked to OKS N-terminal (L1-OKS), as this configuration produced significant lower levels of flavokermesic acid than all other constructs in this sample set (Figure 4.3). The low production of flavokermesic acid could not be explained by higher production of the spontaneous shunt products, octaketides SEK4 and SEK4b, as the total production of metabolites was also significantly lower (Table 4.2). The negative effect of the L1 linker was only observed if it was linked N-terminally onto the OKS. Surprisingly, N-terminal linking of L2 did not have a negative impact on the OKS, though the two linkers were similar. A possible bias in this analysis was that the performance of the different OKS versions was assayed at cell level, normalized to biomass content, rather than the OKS protein levels. Meaning that it was impossible to rule out that the lower level of flavokermesic acid from the strain expressing L1-OKS was due to a lower protein level of L1-OKS as compared to the other strains.

To rule out this possibility we next analyzed the OKS protein levels in the different OKS expressing strains (Table 4.3). This analysis showed that the L1-OKS strain had the highest OKS protein levels of the analyzed strains, with a OKS relative abundance of 113 compared to the average OKS relative abundance of 96 for all strains in this sampleset (FusP-SC01 to FusP-SC06). Proteomics results showed that N-terminal fusion of L1 to the OKS did not reduce the translation rate nor had a negative effect on the half-life of the OKS. Based on this we could therefore concluded that the lower output of flavokermesic acid could be attributed to inefficiency of the L1-OKS enzyme. This could impair OKS when fused with

Table 4.2 Production of flavokermesic acid by strains expressing the OKS enzyme. Numbers in parentheses are standard deviations. Total and flavokermesic acid (FK) concentrations are given in average mg compound per L culture. C7-C12% are the percentage of compound having the C7-C12 fold (SEK4, dSEK4 and flavokeremsic acid). FK% is the percentage of total compound being flavokermesic acid.

		Production of metabolites		Percentage of metabolites	
Sc#	**Enzyme**	**Total**	**FK**	**C7-C12 %**	**FK %**
	Strains expressing OKS only				
Sc01	OKS	0.97 (0.20)	0.24 (0.06)	36.45 (10.6)	28.57 (13.08)
Sc02	OKS-L1	0.64 (0.2)	0.24 (0.04)	44.2 (10.56)	38.89 (11.75)
Sc03	L1-OKS	0.12 (0)	0.08 (0.01)	66.26 (6.37)	66.26 (6.37)
Sc04	OKS-L2	1.17 (0.58)	0.24 (0.07)	33.89 (6.8)	22.47 (8.86)
Sc05	L2-OKS	0.57 (0.04)	0.28 (0.06)	52.71 (6.3)	48.72 (7.75)
Sc06	OKS + XI-2(empty)	1.03 (0.29)	0.22 (0.04)	31.91 (1.2)	21.58 (2.91)
	Strains expressing OKS and one cyclase				
Sc07	OKS + ZhuI	1.21 (0.3)	0.57 (0.13)	72.82 (2.8)	47.42 (5.99)
Sc08	OKS + L1-ZhuI	3.89 (0.1)	0.49 (0.07)	53.33 (0.8)	12.7 (1.4)
Sc09	OKS + ZhuI-L1	4.69 (0.12)	0.55 (0.06)	55.47 (4.36)	11.65 (1.24)
Sc10	OKS + L2-ZhuI	3.94 (0.03)	0.63 (0.05)	56.09 (1.76)	16.02 (1.44)
Sc11	OKS + L1-ZhuI-L2	3.86 (0.27)	0.63 (0.13)	57.57 (2.38)	16.25 (2.4)
Sc12	OKS-L1-ZhuI	0.44 (0.16)	0.3 (0.01)	78.11 (17.02)	74.24 (22.94)
Sc13	ZhuI-L1-OKS	1.37 (0.89)	0.39 (0.05)	55.29 (15.67)	37.57 (23)
Sc14	OKS-L2-ZhuI	0.88 (0.73)	0.24 (0.03)	52.08 (15.58)	38.5 (19.53)
Sc15	ZhuI-L2-OKS	0.93 (0.2)	0.32 (0.04)	50.2 (7.14)	35.96 (11.72)
Sc16	OKS + ZhuJ	12.36 (1.82)	12.09 (1.67)	97.96 (0.85)	97.84 (0.9)
Sc17	OKS + L1-ZhuJ	10.41 (0.83)	7.34 (0.97)	80.96 (1.27)	70.35 (3.91)
Sc18	OKS + ZhuJ-L1	23.63 (0.78)	21.15 (0.82)	92.07 (0.79)	89.5 (0.67)
Sc19	OKS + L2-ZhuJ	10.71 (3.36)	8.26 (3.35)	83.47 (2.98)	75.19 (9.34)
Sc20	OKS + L1-ZhuJ-L2	12.41 (1.47)	9.65 (1.5)	85.73 (2.38)	77.57 (3.07)
Sc21	OKS-L1-ZhuJ	0.53 (0.13)	0.53 (0.13)	99.66 (0.59)	99.66 (0.59)
Sc22	ZhuJ-L1-OKS	0.39 (0.1)	0.39 (0.1)	100 (0)	100 (0)
Sc23	OKS-L2-ZhuJ	0.67 (0.2)	0.67 (0.2)	99.13 (1.33)	99.13 (1.33)
Sc24	ZhuJ-L2-OKS	0.37 (0.03)	0.37 (0.03)	100 (0)	100 (0)
	Strains expressing OKS and two cyclases				
Sc25	OKS + ZhuI + ZhuJ	52.26 (4.16)	52.23 (4.17)	99.95 (0.04)	99.95 (0.04)
Sc26	OKS + ZhuI-L1-ZhuJ	34.58 (16.19)	34.37 (16.07)	99.49 (0.48)	99.43 (0.54)
Sc27	OKS + ZhuI-L2-ZhuJ	24.75 (1.57)	24.51 (1.33)	99.12 (0.86)	99.04 (0.97)
Sc28	OKS + ZhuJ-L1-ZhuI	3.55 (0.58)	3.3 (0.39)	94.46 (4.64)	93.49 (5.74)
Sc29	OKS + ZhuJ-L2-ZhuI	5.28 (0.31)	5.04 (0.12)	96.41 (3.29)	95.72 (4.1)
Sc30	OKS-L1-ZhuI + ZhuJ	38.51 (3.89)	38.44 (3.84)	99.85 (0.13)	99.84 (0.15)
Sc32	OKS-L2-ZhuI + ZhuJ	40.57 (4.01)	40.57 (4.01)	100 (0)	100 (0)
Sc34	OKS-L1-ZhuJ + ZhuI	3.27 (0.82)	3.27 (0.82)	100 (0)	100 (0)
Sc35	ZhuJ-L1-OKS + ZhuI	2.94 (0.84)	2.94 (0.84)	100 (0)	100 (0)
Sc36	OKS-L2-ZhuJ + ZhuI	2.64 (0.32)	2.64 (0.32)	100 (0)	100 (0)
Sc37	ZhuJ-L2-OKS + ZhuI	1.48 (0.77)	1.48 (0.77)	100 (0)	100 (0)
Sc38	OKS-L1-ZhuI-L2-ZhuJ	7.25 (0.36)	7.25 (0.36)	100 (0)	100 (0)
Sc40	ZhuI-L1-OKS-L2-ZhuJ	0.85 (0.21)	0.85 (0.21)	100 (0)	100 (0)
Sc41	ZhuJ-L1-OKS-L2-ZhuI	0.37 (0.13)	0.37 (0.13)	100 (0)	100 (0)
Sc42	ZhuI-L1-ZhuJ-L2-OKS	16.61 (1.42)	14.97 (1.4)	92.11 (0.75)	90.1 (0.78)
Sc43	ZhuJ-L1-ZhuI-L2-OKS	0.69 (0.11)	0.07 (0.05)	31.61 (5.53)	9.88 (6.95)

the L1 linker in the N-terminus to other biosynthetic enzymes.

Fusion of OKS with one cyclase

Next, we wanted to investigate what effect translational fusion of the OKS with one of the cyclases ZhuI or ZhuJ would have on the production of flavokermesic acid and specificity of the pathway. This was investigated by co-expressing the OKS

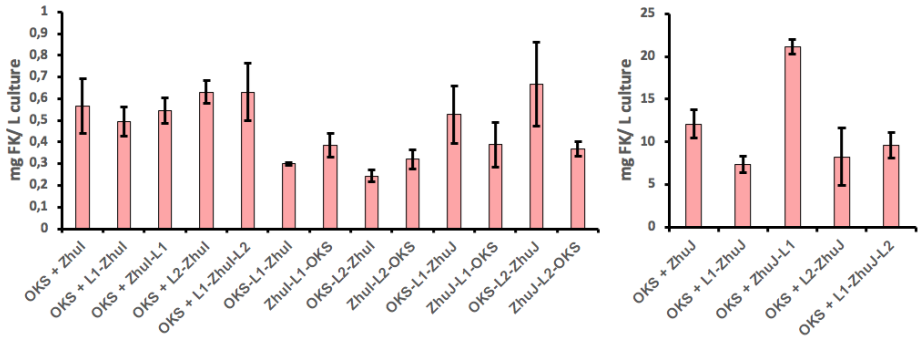


Figure 4.4 Production of Flavokermesic acid by strains expressing the OKS and one of the cyclases ZhuI or ZhuJ. Low producers of flavokermesic acid (<1mg/L culture) are in the left panel while the high producers of flavokermesic acid (>5mg/L culture) are in the right panel.

in combination one of the cyclases as fusion enzymes in all possible configurations (8 strains, FusP-SC12-15 and FusP-SC21-24) and also we wanted to investigate the effect of appending linkers to the cyclases (8 strains, FusP-SC08-11 and FusP-SC17-20); L1 or L2 linker in the N-terminus, the L1 linker in the C-terminus or the L1 linker in the N-terminus and the L2 linker in the C-terminus. This should cover all of the possible configurations in which the cyclases could have a linker. The production of flavokermesic acid in these strains was compared to the flavokermesic acid production in the initial expressions of the OKS and the cyclases individually. The strains were then cultured and the production of flavokermesic acid determined by HPLC-MS (Figure 4.4). Analysis of the results revealed that the 18 strains divided into two groups that differed in the levels of flavokermesic acid produced. A low producing group with flavokermesic levels below 1 mg/L culture consisting of constructs co-expressing OKS and ZhuI and constructs expressing OKS and ZhuJ as a fused enzymes (Figure 4.4, left panel). The second group of strains that all produced flavokermesic acid at levels higher than 5 mg per L culture, consisted of strains co-expressing OKS and ZhuJ as standalone enzymes (Figure 4.4, right panel).

ZhuI is in the pathway responsible for catalyzing formation of the first aromatic ring via a C7-C12 bond (Figure 4.1), direct the metabolic flux toward formation of SEK4 and flavokermesic acid. A fair evaluation of the functionality of ZhuI should therefore not only focus on flavokermesic acid, but instead compare the percentage of metabolites having the C7-C12 fold (C7-C12%) of strains expressing the ZhuI combinations with OKS to that of strains only expressing the OKS

(Figure 4.5). Analysis of the strains showed all strains expressing the OKS and ZhuI as standalone enzymes (FusP-SC07 to FusP-SC11) had a significantly higher C7-C12% than strains only expressing OKS (FusP-SC01), 53-73% compared to 36%. The data here confirms the findings by Andersen-Ranberg *et al.* (2017) [4] and Frandsen *et al.* (2018) [5], in which ZhuI was found to direct synthesis in the desired direction. Evaluation of the effect of the used linkers on the performance of the ZhuI revealed that both linkers had a detrimental effect on the efficiency of ZhuI, when comparing to coexpression of OKS + ZhuI without linkers. The level of flavokermesic acid was found to be significantly higher for all strains co-expressing OKS and ZhuI as standalone enzymes, compared to strains with only the OKS, indicating that when the first ring was formed correctly by ZhuI the spontaneous formation of flavokermesic acid was still possible. When ZhuI and OKS were linked together the C7-C12% was comparable to the standalone expression. Even though ZhuI was still able to fold the nascent octaketide chain when fused to OKS the production of flavokermesic acid was not significantly higher indicating that the OKS was not as efficient when fused to ZhuI. An interesting observation was that the strains expressing fusions of OKS an ZhuI with ZhuI in the N-terminus had a slightly higher average flavokermesic acid level, though not significant, than the strains with ZhuI in the C-terminus (Figure 4.4, left panel).

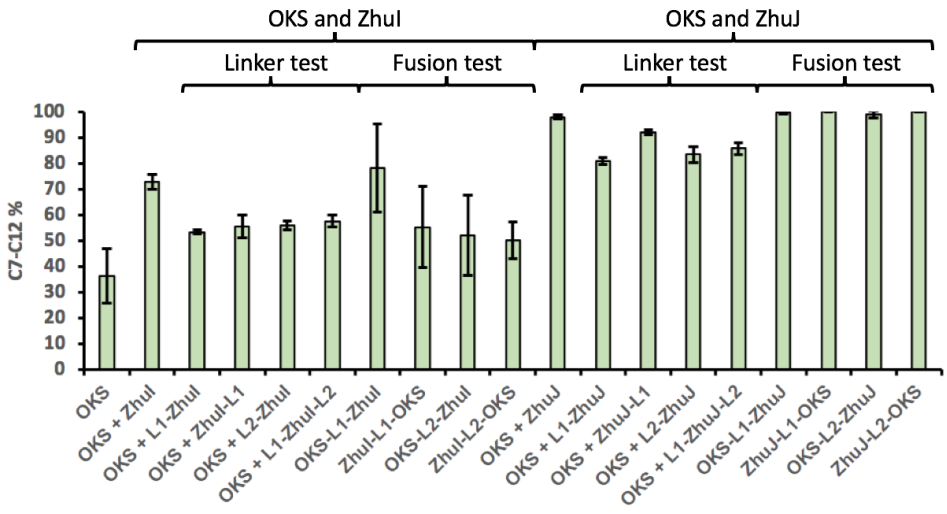


Figure 4.5 Percentage of quantified metabolites having the C7-C12 folding pattern. The effect of linkers on the efficiency of the cyclases as well as effect of fusion is highlighted.

The co-expression of the cyclase ZhuJ and OKS as standalone enzymes led

to a strong increase in the level of flavokermesic acid as previously mentioned. Compared to the strain only expressing the OKS the strains co-expressing the ZhuJ cyclase as a standalone enzyme, including expression of ZhuJ with linkers, led to a 30 to 89 fold increase in flavokermesic acid production. Surprisingly, the strain co-expressing OKS and the ZhuJ-L1 had a significantly ($p < 0.05$) higher production of flavokermesic acid than the other strains expressing the OKS and ZhuJ as standalone enzymes (Appendix Figure S4.2). This indicated that L1 had a positive effect on ZhuJ activity when linked to the C-terminus of the enzyme. Possible explanations for this could be increased substrate uptake or enzymatic turnover rate for the ZhuJ enzyme. An alternative explanation could be faster translation or slower degradation of the enzyme as the protein level of ZhuJ-L1 was considerably higher than the other strains expressing the OKS and ZhuJ as standalone enzymes (FusP-SC16 to FusP-SC20, table 4.3). The constructs expressing the fused versions of OKS and ZhuJ all had significantly higher flavokermesic acid levels compared to the strain expressing OKS except the strain expressing ZhuJ-L1-OKS. With that being said the level of flavokermesic acid was significantly lower than the strains co-expressing OKS and ZhuJ as standalone enzymes. Therefore, it can be concluded that ZhuJ was functional when fused to OKS, but as concluded for OKS and ZhuI fusions OKS was not as productive when fused to ZhuJ. Interestingly both the fused and non-fused ZhuJ and OKS co-expressions had a very low levels of the shunt products SEK4, SEK4b, dSEK4 and dSEK4b with all having above 70% of quantified products (FK%) being flavokermesic acid. The fusion enzymes had a FK% which was close to 100% (Figure 4.5), indicating that when fused to OKS, ZhuJ was very successful in converting the nascent octaketide to flavokermesic acid instead of letting it fold to the shunt products. This could indicate that ZhuJ does not need the C7-C12 ring to form before it can make the C5-C14 ring formation and it may even catalyze the formation of C7-C12 at the same time of making the C5-C14 ring. Inversely to what was found for ZhuI the strains expressing ZhuJ in the C-terminus had a slightly higher average level of flavokermesic acid than the strains expressing ZhuJ in the N-terminus.

To conclude on the productivity of the OKS in combination with one of the cyclases, the co-expression of ZhuI and OKS had a significant effect on the C7-C12%, but not a significant effect on the level of flavokermesic acid for the non-fused versions of the co-expression. Co-expression of ZhuJ and OKS led to elevated levels of flavokermesic acid. This effect was only observed for non-fused versions. Thus,

for both the ZhuI and ZhuJ the fusion with OKS did not have a positive effect on the flux towards flavokermesic acid compared to the non-fused versions. The highest production of flavokermesic acid was with co-expression of OKS and ZhuJ-L1 which had a 89-fold increase in the flavokermesic acid production compared to the strain expressing only OKS.

Fusion of the OKS with both cyclases

The final sample set consisted of strains co-expressing OKS, ZhuI and ZhuJ and fusions of these in different sequences. The expectations were that we would reach the highest levels of flavokermesic acid as this was found in previous studies [5]. Similarly to the strains expressing two enzymes these strains could be split into two different groups based on their production of flavokermesic acid. A low producing group produced up to 5 mg flavokermesic acid per L culture (Figure 4.6, left panel) while the high producing group reached level in the range from 7 to 52 mg flavokermesic acid per L culture (Figure 4.6, right panel). Unfortunately, three of the constructed strains, namely ZhuI-L1-OKS + ZhuJ, ZhuI-L2-OKS + ZhuJ and OKS-L1-ZhuJ-L2-ZhuI, did produce detectable levels of the metabolites quantified here. For the strains expressing ZhuI-L1-OKS + ZhuJ and ZhuI-L2-OKS + ZhuJ this was very surprising as both parent strains (ZhuI-L1-OKS and ZhuI-L2-OKS) produced metabolites before the introduction of the ZhuJ encoding gene. The protein levels were also detected lower for these strains indicating issues with protein expression (Table 4.3). This situation points towards the introduction of critical mutation(s) in the OKS encoding gene during or after the second transformation step.

The strain expressing ZhuI, ZhuJ and OKS as standalone enzymes (FusP-SC25) had an average level of 52 mg flavokermesic acid per L culture, as previously mentioned, and this was also the highest level detected for all created strains. As good as all of the quantified metabolites were detected as flavokermesic acid (FK% = 99,95, table 4.2) and the flavokermesic acid level was 1.3 fold higher than second highest producer expressing OKS-L2-ZhuI + ZhuJ. This shows that not only was the cell factory very efficient at producing flavokermesic acid, but at the same time generated very few shunt products were formed.

The strains co-expressing OKS as a standalone enzyme with ZhuI and ZhuJ fused together showed that a N-terminal localization of ZhuI resulted in 22 to 25 mg flavokermesic acid per L culture compared to only 4 to 5 mg flavokermesic

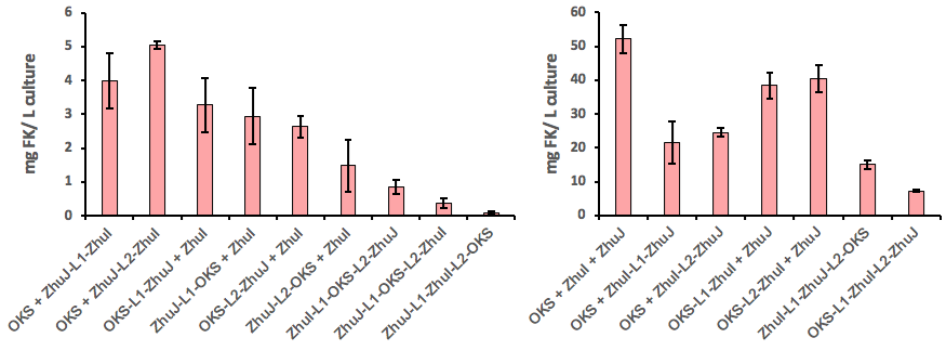


Figure 4.6 Production of Flavokermesic acid by strains expressing the OKS and both of the cyclases ZhuI and ZhuJ.

acid per L culture for strains having the opposite sequence of ZhuI and ZhuJ. This finding was in agreement with the finding from expressing fusion enzymes of the OKS and ZhuI or ZhuJ, which also found that the level of flavokermesic acid was highest with ZhuI in the N-terminus and ZhuJ in the C-terminus.

The strains expressing fusion enzymes of OKS and ZhuI or ZhuJ likewise showed different levels of flavokermesic acid when supplemented with the missing cyclase as a standalone enzyme. All strains except the strain expressing ZhuJ-L2-OKS + ZhuI had significantly higher levels of flavokermesic acid than the strains only expressing the fused version without the standalone supplementary cyclase. This shows that the addition of the standalone cyclase drives the overall production of flavokermesic acid. The strains expressing OKS-L1-ZhuI + ZhuJ and OKS-L2-ZhuI + ZhuJ reached very high levels of flavokermesic acid of 38 to 41 mg flavokermesic acid per L culture. It was very unfortunate that the strains expressing ZhuI-L1-OKS + ZhuJ and ZhuI-L2-OKS + ZhuJ did not produce any detectable metabolites as these had a higher level of flavokermesic acid in the previous dataset with only OKS and one cyclase. Therefore, it was expected that these strains would have a higher production level of flavokermesic acid, which could even be as high or higher than the level reached by the standalone expression of the three biosynthetic enzymes. The strains expressing the fusion of OKS and the ZhuJ cyclase with standalone expression of ZhuI all had comparable levels of flavokermesic acid between 1 and 3 mg flavokermesic acid per L culture. This level was significantly higher than the strains expressing the OKS and ZhuJ fusions without the standalone expression of ZhuI except for ZhuJ-L2-OKS + ZhuI as mentioned earlier, but the level of flavokermesic acid produced was still low

compared to other constructs expressing all three biosynthetic enzymes.

Finally, for strains expressing the OKS and the two cyclases as one fusion enzyme, the highest average levels of flavokermesic acid was obtained by strains expressing OKS-L1-ZhuI-L2-ZhuJ and ZhuI-L1-ZhuJ-L2-OKS producing 7 and 15 mg flavokermesic acid per L culture respectively. These strains each partly follow the tendency of having ZhuI in the N-terminus and ZhuJ in the C-terminus for having high production of flavokermesic acid. Interestingly the two strains expressing OKS in the C-terminus of the fusion enzymes with either of the cyclases had very different performances, where the ZhuI-L1-ZhuJ-L2-OKS expressing strain produced 15 mg flavokermesic acid per L culture with a C7-C12% of 92 while the ZhuJ-L1-ZhuI-L2-OKS expressing strain produced 0.1 mg flavokermesic acid per L culture with a C7-C12% of 32. This significant difference reveals that the neither ZhuJ nor ZhuI was very efficient in the latter construct. The remaining strains expressed ZhuI-L1-OKS-L2-ZhuJ and ZhuJ-L1-OKS-L2-ZhuI and both were not very successful in producing flavokermesic acid with levels of 0.8 and 0.4 mg flavokermesic acid per L culture respectively. This indicated that having the OKS as the middle enzyme was very unfavorable.

To conclude on expression of OKS and both of the cyclases it was found that the standalone expression of all enzymes reached the highest production level of flavokermesic acid. The tendency found in fusion of the OKS with one of the cyclases showed that ZhuI performs best in the N-terminus and ZhuJ in the C-terminus was confirmed in this dataset as constructs abiding by this rule reached significantly higher production levels of flavokermesic acid. The strains expressing fusions of all three enzymes were not as successful in producing flavokermesic acid as other constructs expressing the enzymes as fusions of two of the enzymes or as standalone enzymes.

4.4.3 Protein quantification

The proteomics dataset has been used to explain performance of individual strains in the metabolomics dataset. In this section focus will be put on finding general trends of how the protein levels did or did not vary between different groups of constructs. A number of calculations were done to compare different constructs and an overview of these along with the detected protein abundances for individual strains can be found in table S4.6 (Whole proteomics dataset and calculated values are available in appendix table 4.3).

Table 4.3 Overview of the proteomics dataset collected in this study. Protein levels are relative to the levels recorded in the normalization sample. Full dataset is available in appendix table S4.6

Strain #	Enzymes	OKS	ZhuI	ZhuJ
FusP-SC01	OKS	87.1	-	-
FusP-SC02	OKS-L1	75.4	-	-
FusP-SC03	L1-OKS	113.3	-	-
FusP-SC04	OKS-L2	90.8	-	-
FusP-SC05	L2-OKS	105.2	-	-
FusP-SC06	OKS + XI-2(empty)	101.6	-	-
FusP-SC07	OKS + ZhuI	88.9	202.9	-
FusP-SC08	OKS + L1-ZhuI	92.3	174.1	-
FusP-SC09	OKS + ZhuI-L1	79.1	360.1	-
FusP-SC10	OKS + L2-ZhuI	88.6	152.5	-
FusP-SC11	OKS + L1-ZhuI-L2	75.4	164.7	-
FusP-SC12	OKS-L1-ZhuI	92.8	78.3	-
FusP-SC13	ZhuI-L1-OKS	127	118.8	-
FusP-SC14	OKS-L2-ZhuI	78.7	115.5	-
FusP-SC15	ZhuI-L2-OKS	103.4	96	-
FusP-SC16	OKS + ZhuJ	95.9	-	312
FusP-SC17	OKS + L1-ZhuJ	95.8	-	108.5
FusP-SC18	OKS + ZhuJ-L1	98.2	-	370.4
FusP-SC19	OKS + L2-ZhuJ	89.3	-	158.8
FusP-SC20	OKS + L1-ZhuJ-L2	90.8	-	168.7
FusP-SC21	OKS-L1-ZhuJ	37.2	-	137.4
FusP-SC22	ZhuJ-L1-OKS	48.5	-	82.4
FusP-SC23	OKS-L2-ZhuJ	47.4	-	106
FusP-SC24	ZhuJ-L2-OKS	38.2	-	85.6
FusP-SC25	OKS + ZhuI + ZhuJ	100.6	207.5	175
FusP-SC26	OKS + ZhuI-L1-ZhuJ	91.4	55.4	214.8
FusP-SC27	OKS + ZhuI-L2-ZhuJ	83.9	64.6	220.1
FusP-SC28	OKS + ZhuJ-L1-ZhuI	81.7	31	89.8
FusP-SC29	OKS + ZhuJ-L2-ZhuI	93.5	54.8	132.5
FusP-SC30	OKS-L1-ZhuI + ZhuJ	61.8	83.1	476.4
FusP-SC32	OKS-L2-ZhuI + ZhuJ	84.1	91.4	386.7
FusP-SC34	OKS-L1-ZhuJ + ZhuI	51.8	224.9	122.5
FusP-SC35	ZhuJ-L1-OKS + ZhuI	32.9	277.3	84.9
FusP-SC36	OKS-L2-ZhuJ + ZhuI	47.5	166.3	141.3
FusP-SC37	ZhuJ-L2-OKS + ZhuI	40.4	214.6	130.4
FusP-SC38	OKS-L1-ZhuI-L2-ZhuJ	44.9	35.8	121.8
FusP-SC40	ZhuI-L1-OKS-L2-ZhuJ	36.1	60.2	95.8
FusP-SC41	ZhuJ-L1-OKS-L2-ZhuI	43.1	46.3	134.7
FusP-SC42	ZhuI-L1-ZhuJ-L2-OKS	46.5	66.4	172.6
FusP-SC43	ZhuJ-L1-ZhuI-L2-OKS	43.5	28.4	88.3

Initially, we wanted to investigate how fusions affected the enzyme levels of the individual biosynthetic enzymes. To allow for this comparison we calculated the average enzyme levels for each of the enzymes OKS, ZhuI and ZhuJ, in strains expressing these as standalone enzymes or as fusion enzymes with one or two of the other enzymes. This calculation revealed that the levels of enzymes were significantly higher for enzymes expressed as standalone enzymes compared to the same enzymes expressed in fusions (Table 4.4). The levels of enzymes, comparing fused enzymes, were higher when two enzymes were fused than when three were fused though this was not significant. Therefore, we concluded that fusion of the enzymes had a detrimental effect on the overall level of enzyme available for

biosynthesis and thereby also the possible output of the artificial biosynthetic pathway. This lower level of enzyme when expressed as fusion enzymes could be due to issues with translation of the larger enzymes or incomplete folding as a consequence of the fusion and therefore higher degradation rates. Alternatively, issues could be linked to secondary structures in the mRNA, which are only present in fusion of two or three of the enzymes. These secondary structures could impair translation and thus result in a lower level of protein.

As mentioned earlier in the metabolomics section, strains expressing fused version of enzymes had a tendency for higher flavokermesic acid production if ZhuI was placed in the N-terminus and ZhuJ in the C-terminus. This lead to the question of whether the higher productivity was due to higher levels of enzyme in these strains or if it was truly due to the two enzymes being more effective in this sequence. To settle this we analyzed the data for the strains expressing fusion enzymes consisting of the OKS and one of the two cyclases. The enzyme levels of these fusion enzymes were not significantly different from one another (Appendix table S4.6) and therefore, it can be concluded that the higher output of enzymes having ZhuI in the N-terminus and/or ZhuJ in the C-terminus can be coupled to the fusion enzymes catalytic properties rather than the enzyme levels in the cell.

Next, we wanted to investigate if and how high levels of flavokermesic acid were correlated to the level of the individual enzymes. For this we compared the enzyme levels in a group of high and low performing strains that expressed all three biosynthetic enzymes, either as fusions or standalone enzymes. The high performers consisted of 5 strains that all produced more than 10 mg flavokermesic acid per L culture, while the 5 strains of low performers produced less than 3 mg flavokermesic acid per L culture. The comparison (Table 4.5) showed that enzyme levels for OKS and ZhuJ along with the summed level of OKS, ZhuI and ZhuJ proteins were all significantly higher for strains producing high levels of flavokermesic acid, and that the enzyme level for ZhuI was lower, though insignificant, for the strains producing high levels of flavokermesic acid. The same tendency was found

Table 4.4 The level of enzymes when expressed as standalone enzymes or dual or triple fusion enzymes. Significance (T-test) between the different groups were calculated. (Sin - single, Dua - Dual fusion, Tri - Triple fusion)

	Enzymatic levels			**Significance**		
Enz	**Sin**	**Dua**	**Tri**	**Sin vs Dua**	**Sin vs Tri**	**Dua vs Tri**
OKS	91.4 (9.5)	61.3 (27.7)	43.9 (4.4)	4.52 E-05	7.49 E-12	0.15
ZhuI	214.5 (63.0)	71.9 (30.5)	49.2 (15.0)	1.41 E-07	1.11 E-06	0.11
ZhuJ	281.6 (124.5)	129.0 (46.7)	121.2 (30.3)	2.72 E-05	8.08 E-04	0.72

Table 4.5 Comparison of a strains expressing all three biosynthetic enzymes divided into a high producing group of strains and a low producing group of strains with regards to the level of flavokermesic acid.

Enz.	High producers	Low producers	T-test HP vs LP
OKS	78.1 (20.1)	40.6 (5.3)	0.001
ZhuI	94.7 (56.8)	132.2 (102.3)	0.451
ZhuJ	274.3 (126.6)	112.6 (25.6)	0.012
Summed	447.1 (130.0)	285.3 (104.9)	0.039

for strains expressing the OKS and one of the cyclases (Table 4.3). Therefore, it appears that high levels of flavokermesic acid can be reached independently from the concentration of ZhuI. This could mean that ZhuI was extremely efficient at folding the nascent octaketide produced by OKS. This hypothesis was however not supported by the data for the strains that co-expressed OKS and ZhuI as standalone or as fusion enzymes without ZhuJ as all had a C7-C12% between 50 and 80, compared to the co-expression of OKS and ZhuJ having a C7-C12% between 80 to 100. With that being said the co-expression of all three enzymes yielded the best C7-C12% between 92 to 100 with the exception of ZhuJ-L1-ZhuI-L2-OKS having a C7-C12% of 32 (Table 4.2).

Therefore, it can be concluded that ZhuI had a positive effect when co-expressed with OKS and ZhuJ while the effect of ZhuI was not as clear when only co-expressed with OKS. The pattern we found was different to findings in Frandsen *et al.* (2018) where the co-expression of OKS and ZhuI in *A. nidulans* lead to a 2.1-fold increase in flavokermesic acid level and co-expression of OKS, ZhuI and ZhuJ lead to a 4.1 fold increase in flavokermesic acid level. When making the same comparison in our dataset we saw a comparable 2.4-fold increase in flavokermesic acid level when co-expressing OKS and ZhuI compared to expression of OKS alone, while we saw a 220-fold increase when co-expressing ZhuI and ZhuJ with OKS compared to expressing OKS alone. This effect can be attributed to the effect of ZhuJ in *S. cerevisiae* as we saw a 51-fold increase in flavokermesic acid level for co-expression of OKS and ZhuJ compared to expression of OKS, whereas only a 2.7-fold increase was observed by Frandsen *et al.* (2018) [5].

In an attempt to identify the best enzyme configuration for production of flavokermesic acid, we analyzed how enzyme level was correlated to flavokermesic acid level, by normalizing the flavokermesic acid levels to the level of the three enzymes for each construct. This was calculated as flavokermesic acid production

divided by the enzyme levels of OKS, ZhuI, ZhuJ or the summed amount of OKS, ZhuI and ZhuJ enzymes in each sample (Table 4.6). This analysis revealed that strains that produced the highest levels of flavokermesic acid also had the highest ratios of flavokermesic acid per enzyme, showing that high flavokermesic acid levels were not only due to a higher concentration of the enzymes, but also a function of the nature of the constructs. This support the claim that the enzyme configurations impact the efficiency of the pathway towards flavokermesic acid. The highest flavokermesic acid per OKS ratio was observed in the strain expressing OKS-L1-ZhuI + ZhuJ, closely followed by the highest flavokermesic acid producer expressing OKS + ZhuI + ZhuJ. The highest enzyme specific activity (mg flavokermesic acid per summed enzyme level) was obtained by the strain expressing OKS + ZhuI + ZhuJ.

From the proteomics dataset it can be concluded that the higher levels of flavokermesic acid observed for the cell factory was not only depending on actual enzyme levels, but that different constructs also featured different performance levels with regards to their ability to produce flavokermesic acid.

4.4.4 Potential issues with fusion of biosynthetic enzymes

The second aim of this study was to investigate if fusing the enzymes in the artificial pathway producing flavokermesic acid, could improve the overall productivity and specificity of the cell factory. The hope was that the physical linkage of the enzymes would increase the local substrate levels for the individual enzymes and thereby reducing the time the highly reactive intermediates would spend in the free form before encountering the next enzyme in the pathway. This rational has in several systems been proven correct, e.g. production of terpenes or small aromatic compounds [8, 18]. However, this was unfortunately not true for the pathway in question as translational fusion of the enzymes had detrimental effects on the performance of the pathway. A possible explanation for this may be that the three enzymes in this study do not act as monomers in solution in their natural systems, but form homodimers or -multimers to achieve functionality. Fusion with the other biosynthetic enzymes could therefore hinder the formation of these dimers.

The structure of OKS was elucidated by Morita *et al.* (2007) [19]. The structure has not been made public but was found to be very homologous to that of the pentaketide chromone synthase for which the structure is public. OKS is a dimer which could lead to structural clashes and problem when attempting to make fu-

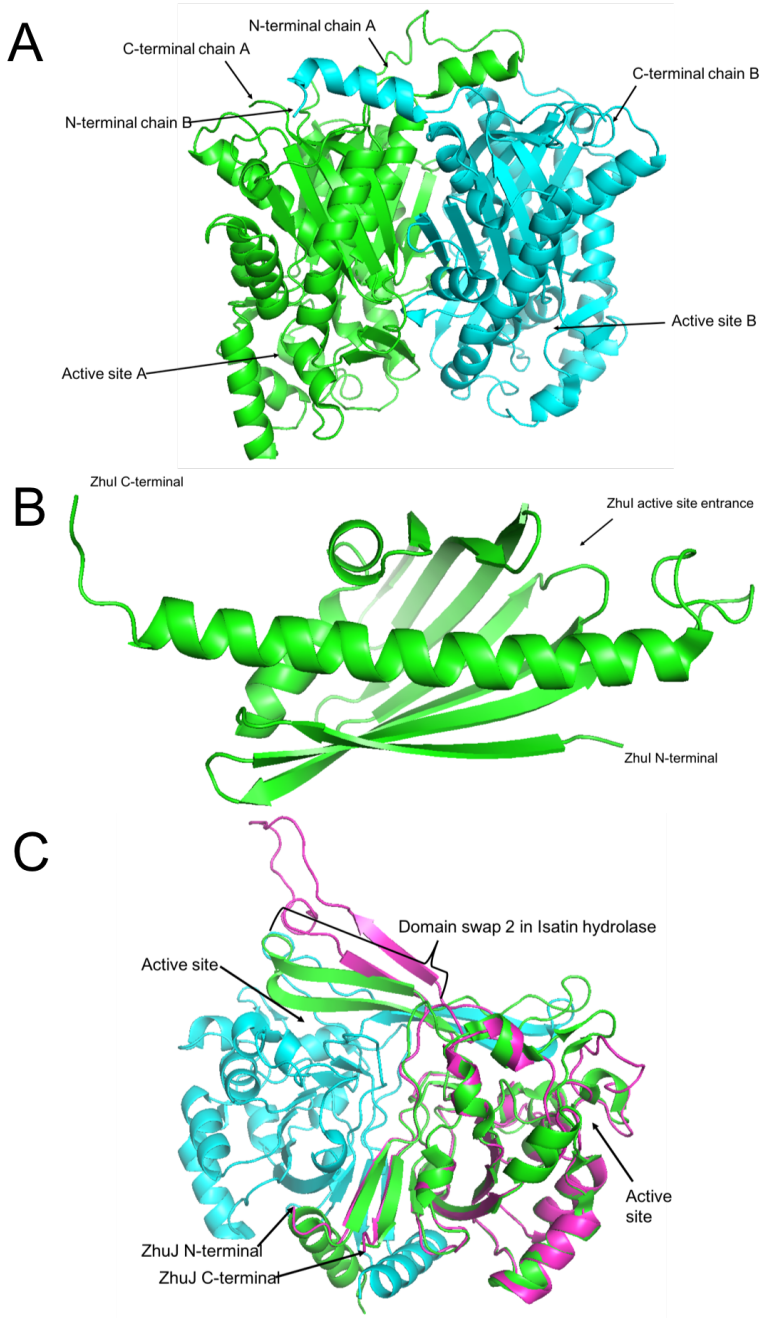


Figure 4.7 Protein structures of the Biosynthetic enzymes. (A) Structure of PCS (PDB code: 2D52) with N- and C-termini of the dimer highlighted. (B) Structure of ZhuI (PDB code: 3TFZ) with N- and C-termini, and the active site entrance of the dimer highlighted. (C) Modelled structure of ZhuJ superimposed on the Isatin hydrolase (PDB: 4M8D) which was used as the template. ZhuJ is shown in magenta and the Isatin hydrolase dimer is shown in cyan and green. N- and C-termini of the dimer highlighted.

Table 4.6 Calculated values comparing the level of flavokermesic acid and the enzymatic levels of OKS, ZhuI and ZhuJ individually and also the summed amount of OKS, ZhuI and ZhuJ enzymes for each strain. The list is sorted by the level of flavokermesic acid produced by the individual strains (Conc. FK).

Strain	**Conc. FK**	**FK/OKS**	**FK/ZhuI**	**FK/ZhuJ**	**FK/summed**
OKS + ZhuI + ZhuJ	52.26	0.519	0.252	0.299	108.171
OKS-L2-ZhuI + ZhuJ	40.57	0.482	0.444	0.105	72.165
OKS-L1-ZhuI + ZhuJ	38.51	0.623	0.463	0.081	61.982
OKS + ZhuI-L1-ZhuJ	34.58	0.378	0.624	0.161	95.631
OKS + ZhuI-L2-ZhuJ	24.75	0.295	0.383	0.112	67.151
OKS + ZhuJ-L1	23.63	0.241	-	0.064	45.228
ZhuI-L1-ZhuJ-L2-OKS	16.61	0.357	0.250	0.096	58.177
OKS + L1-ZhuJ-L2	12.41	0.137	-	0.074	41.180
OKS + ZhuJ	12.36	0.129	-	0.040	27.109
OKS + L2-ZhuJ	10.71	0.120	-	0.067	34.716
OKS + L1-ZhuJ	10.41	0.109	-	0.096	45.068
OKS-L1-ZhuI-L2-ZhuJ	7.25	0.161	0.202	0.060	35.790
OKS + ZhuJ-L2-ZhuI	5.28	0.056	0.096	0.040	18.797
OKS + ZhuI-L1	4.69	0.059	0.013	-	9.247
OKS + L2-ZhuI	3.94	0.044	0.026	-	10.698
OKS + L1-ZhuI	3.89	0.042	0.022	-	10.539
OKS + L1-ZhuI-L2	3.86	0.051	0.023	-	10.907
OKS + ZhuJ-L1-ZhuI	3.55	0.043	0.115	0.040	17.546
OKS-L1-ZhuJ + ZhuI	3.27	0.063	0.015	0.027	8.182
ZhuJ-L1-OKS + ZhuI	2.94	0.089	0.011	0.035	7.441
OKS-L2-ZhuJ + ZhuI	2.64	0.056	0.016	0.019	7.427
ZhuJ-L2-OKS + ZhuI	1.48	0.037	0.007	0.011	3.849
ZhuI-L1-OKS	1.37	0.011	0.012	-	3.644
OKS + ZhuI	1.21	0.014	0.006	-	3.283
OKS-L2	1.17	0.013	-	-	3.699
OKS + XI-2(empty)	1.03	0.010	-	-	3.273
ZhuI-L2-OKS	0.93	0.009	0.010	-	2.766
OKS-L2-ZhuI	0.88	0.011	0.008	-	3.177
ZhuI-L1-OKS-L2-ZhuJ	0.85	0.023	0.014	0.009	4.410
ZhuJ-L1-ZhuI-L2-OKS	0.69	0.016	0.024	0.008	4.292
OKS-L2-ZhuJ	0.67	0.014	-	0.006	2.975
OKS-L1	0.64	0.008	-	-	2.954
L2-OKS	0.57	0.005	-	-	2.767
OKS-L1-ZhuJ	0.53	0.014	-	0.004	2.155
OKS	0.44	0.005	-	-	1.797
OKS-L1-ZhuI	0.44	0.005	0.006	-	1.790
ZhuJ-L1-OKS	0.39	0.008	-	0.005	2.481
ZhuJ-L2-OKS	0.37	0.010	-	0.004	2.091
ZhuJ-L1-OKS-L2-ZhuI	0.37	0.008	0.008	0.003	1.634
L1-OKS	0.12	0.001	-	-	0.374

sions with other enzymes as the OKS is only active when the dimer is formed due to the opposite dimer participating in forming the catalytic cavity [20]. Also, the N- and C-termini of the OKS are in close vicinity on the surface, which could lead to issues when attaching linkers and enzymes on both of the termini if the linkers are not sufficiently long to prevent strain on the protein structures due to steric clashes of the enzymes (Figure 4.7).

The structure of ZhuI was elucidated by Ames *et al.* (2011). It is dimeric and features a helix-grip fold [21], which has also been observed for other polyketide cyclases [22, 23]. The N- and C-termini of ZhuI are in opposite ends of the protein which should lead to less issues when attaching enzymes to both ends of the

amino acid chain. That being said, the trend for ZhuI to maximize conversion to flavokermesic acid was that ZhuI should be in the N-terminal end of the fusion protein. This could be because the N-terminal end of ZhuI was susceptible to strain from the other enzymes in the fused enzyme rendering ZhuI unfunctional or that the attached enzyme could cover the active site entrance of ZhuI (Figure 4.7). The latter is unlikely as the entrance to the active site, though in the N-terminal end of the enzyme, is opposite the central alpha helix (named αC in Ames *et al.* (2011)) and at a distance of 22Å (measured in PyMOL).

The structure of ZhuJ is unknown and no structures exists for closely related cyclases that has been shown to catalyze a similar polyketide folding pattern. The closest amino acid sequence homolog (43% similarity) is the isatin hydrolase *Labrenzia aggregata* (PDB entry 4M8D), which is a dimer and has a swiveling $\alpha/\beta/\alpha$ domain with two essential domains to stabilize the dimer, named domain swaps, as these domains "hug" the other monomer [24]. Based on the sequence of ZhuJ one of these domain swaps is likely present in ZhuJ (Figure 4.7). Other sequence homologous proteins for which a structure is known are kyrnurenine formamidase from *Pseudomonas aeruginosa*, metal-dependent hydrolase from *Bacillus stearothermophilus*, the novel cyclase from an uncultured organism, and protein of unknown function from *Methanocaldococcus jannaschii* (PDB entries 4COB, 1R61, 5IBZ, and 2B0A respectively). These sequence homologs are also dimers and feature the domain swap 2 described in the isatin hydrolase structure. Therefore, we find that it is safe to assume that ZhuJ is also a dimer in solution.

To conclude all of the expressed enzymes are expected to be dimers in solution. The formation of dimers may be troublesome when the enzymes are expressed as fusions either due to direct sterical clashes between the proteins or due to the individual enzymes being unable to form its homodimer. As an example, in the fusion of all three biosynthetic enzymes. The enzymes are assumed to be expressed correctly, but when one of them forms the homodimer the strain on the other enzymes renders the other enzymes in the fusion inactive. Also, if the enzymes are required to be dimeric to be stable, they could therefore denature as a consequence of being monomeric, and therefore be tagged for degradation by the cell. This could offer an explanation to why the enzyme fusions attempted here were not successful in improving the production of flavokermesic acid compared to unfused enzymes. This explanation is also supported by the lower level of enzyme detected when the enzymes were expressed as fusions of two or three enzymes

compared to the expression of the standalone enzymes.

4.5 Conclusion and perspectives

The initial goal of this study was to establish the pathway producing flavokermesic acid in *S. cerevisiae*, which had previously been established in *N. benthamiana* and *A. nidulans*. This was successfully done and the expected metabolites SEK4, SEK4b, dSEK4, dSEK4b and flavokermesic acid were all detected when expressing the artificial pathway in *S. cerevisiae*. Another aim of this study was to investigate if fusion of the three biosynthetic enzymes would yield higher titers of the desired flavokermesic acid and less of the shunt products SEK4, SEK4b, dSEK4 and dSEK4b. The production of shunt products was found to be strongly correlated to which, and how many, of the biosynthetic enzymes were expressed, instead of the option of having the enzymes expressed as fusion or as standalone enzymes. The highest level of 52 mg flavokermesic acid per L of culture was obtained by the strain expressing the three biosynthetic enzymes as standalone enzymes and this level of flavokermesic acid provides a good starting point for creating a cell factory for production of flavokermesic acid and later carminic acid.

Acknowledgements

We would like to thank the DTU Bioengineering Metabolomics Core for assistance in performing the HPLC-MS analysis, and the DTU Bioengineering Proteomics Core for assistance in protein extraction and analysis of proteomes. This work was funded by the Novo Nordisk foundation grant number NNF15OC0016626.

Author contributions

The research study was conceived by RF and KK. The molecular biology work was designed and conducted by KK, CS, TH, RR and AV. The cultivations, chemical analysis and proteomics was done by KK. Drafting of the manuscript was done by KK, and refined by RF and TL.

References

[1] Donkin, RA. "The insect dyes of western and west-central Asia". In: *Anthropos* 5/6 (1977), pp. 847–880.

[2] Baranyovits, FLC. "Cochineal carmine: an ancient dye with a modern role". In: *Endeavour* 2.2 (1978), pp. 85–92.

[3] Rasmussen, SA, Kongstad, KT, Khorsand-Jamal, P, Kannangara, RM, Nafisi, M, van Dam, A, Bennedsen, M, Madsen, B, Okkels, F, Gotfredsen, CH, Stark, D, Thrane, U, Mortensen, UH, Larsen, TO, and Frandsen, RJN. "On the biosynthetic origin of carminic acid". In: *Insect Biochem Mol Biol* 96 (2018), pp. 51–61.

[4] Andersen-Ranberg, J, Kongstad, KT, Nafisi, M, Staerk, D, Okkels, FT, Mortensen, UH, Lindberg Møller, B, Frandsen, RJN, and Kannangara, R. "Synthesis of C-Glucosylated Octaketide Anthraquinones in *Nicotiana benthamiana* by Using a Multispecies-Based Biosynthetic Pathway". In: *ChemBioChem* 18.19 (2017), pp. 1893–1897.

[5] Frandsen, RJ, Khorsand-Jamal, P, Kongstad, KT, Nafisi, M, Kannangara, RM, Staerk, D, Okkels, FT, Binderup, K, Madsen, B, Møller, BL, Thrane, U, and Mortensen, UH. "Heterologous production of the widely used natural food colorant carminic acid in *Aspergillus nidulans*". In: *Sci Rep* 8.12853 (2018), pp. 1–10.

[6] Abe, I, Oguro, S, Utsumi, Y, Sano, Y, and Noguchi, H. "Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase". In: *J Am Chem Soc* 127.36 (2005), pp. 12709–12716.

[7] Marti, T, Hu, Z, Pohl, NL, Shah, AN, and Khosla, C. "Cloning, nucleotide sequence, and heterologous expression of the biosynthetic gene cluster for R1128, a non-steroidal estrogen receptor antagonist insights into an unusual priming mechanism". In: *J Biol Chem* 275.43 (2000), pp. 33443–33448.

[8] Albertsen, L, Chen, Y, Bach, LS, Rattleff, S, Maury, J, Brix, S, Nielsen, J, and Mortensen, UH. "Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes". In: *Appl Environ Microbiol* 77.3 (2011), pp. 1033–1040.

[9] Mikkelsen, MD, Buron, LD, Salomonsen, B, Olsen, CE, Hansen, BG, Mortensen, UH, and Halkier, BA. "Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform". In: *Metab Eng* 14.2 (2012), pp. 104–111.

[10] Sherman, F, Fink, GR, and Hicks, JB. *Laboratory Course Manual for Methods in Yeast Dynamics*. Cold Spring Harbor Laboratory, 1986.

[11] Partow, S, Siewers, V, Nielsen, J, and Maury, J. "Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*". In: *Yeast* 26.10 (2009), pp. 545–551.

[12] Gietz, RD and Woods, RA. "Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method". In: *Method Enzymol* 350 (2002), pp. 87–96.

[13] Kulak, NA, Pichler, G, Paron, I, Nagaraj, N, and Mann, M. "Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells". In: *Nat Meth* 11.3 (2014), p. 319.

[14] Rappsilber, J, Mann, M, and Ishihama, Y. "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips". In: *Nat Protoc* 2.8 (2007), p. 1896.

[15] Schrödinger, LLC. "The PyMOL Molecular Graphics System, Version 1.8". 2015.

[16] Zimmermann, L, Stephens, A, Nam, SZ, Rau, D, Kübler, J, Lozajic, M, Gabler, F, Söding, J, Lupas, AN, and Alva, V. "A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core". In: *J Mol Biol* 430.15 (2018), pp. 2237–2243.

[17] Chen, Y, Daviet, L, Schalk, M, Siewers, V, and Nielsen, J. "Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism". In: *Met Eng* 15 (2013), pp. 48–54.

[18] Lee, D, Lloyd, NDR, Pretorius, IS, and Borneman, AR. "Heterologous production of raspberry ketone in the wine yeast *Saccharomyces cerevisiae* via pathway engineering and synthetic enzyme fusion". In: *Microb Cell Fact* 15.49 (2016), pp. 1–7.

[19] Morita, H, Kondo, S, Kato, R, Wanibuchi, K, Noguchi, H, Sugio, S, Abe, I, and Kohno, T. "Crystallization and preliminary crystallographic analysis of an octaketide-producing plant type III polyketide synthase". In: *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* 63.11 (2007), pp. 947–949.

[20] Abe, I, Utsumi, Y, Oguro, S, Morita, H, Sano, Y, and Noguchi, H. "A plant type III polyketide synthase that produces pentaketide chromone". In: *J Am Chem Soc* 127.5 (2005), pp. 1362–1363.

[21] Ames, BD, Lee, MYY, Moody, C, Zhang, W, Tang, Y, and Tsai, SCC. "Structural and biochemical characterization of ZhuI aromatase/cyclase from the R1128 polyketide pathway". In: *Biochem* 50.39 (2011), pp. 8392–8406.

[22] Lee, MY, Ames, BD, and Tsai, SC. "Insight into the molecular basis of aromatic polyketide cyclization: crystal structure and in vitro characterization of WhiE-ORFVI". In: *Biochem* 51.14 (2012), pp. 3079–3091.

[23] Ames, BD, Korman, TP, Zhang, W, Smith, P, Vu, T, Tang, Y, and Tsai, SC. "Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides". In: *Proc Nat Acad Sci* 105.14 (2008), pp. 5349–5354.

[24] Bjerregaard-Andersen, K, Sommer, T, Jensen, JK, Jochimsen, B, Etzerodt, M, and Morth, JP. "A proton wire and water channel revealed in the crystal structure of isatin hydrolase". In: *J Biol Chem* 289.31 (2014), pp. 21351–21359.

5 Manuscript III – The *pgl1* biosynthetic gene cluster and metabiosynthetic pathway for fusarubins in *Fusarium*

Kromphardt KJK[a], Kim H-S[b], Proctor RH[b], Larsen TO[a] & Frandsen RJN[a]
[a] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark; [b] National Center for Agricultural Utilization Research, US Department of Agriculture, Illinois, USA

This manuscript is planned for submission to Fungal Genetics and Biology.

5.1 Abstract

Species of the fungal genus *Fusarium* have long been in the interest of researches, and because of the toxin production found in several fusaria the metabolic potential has received much attention. This has led to understanding of many of the secondary metabolic pathways for important toxins and pigments in *Fusarium* spp. Here we wanted to investigated the *pgl1* biosynthetic gene cluster across the whole genus of *Fusarium* as well as propose a common biosynthetic pathway for pigments thought to be produced by enzymes from the *pgl1* gene cluster. We have named this the metabiosynthetic pathway. We compared the *pgl1* gene cluster in 22 *Fusarium* species and a synteny plot revealed high conservation of cluster topology throughout the genus of *Fusairum*. The proposed metabiosynthetic pathway showed many common biosynthetic steps, but also highlighted that many biosynthetic steps could not be explained with the current biosynthetic cluster. Therefore, on the basis of a different gene cluster topology and apparent higher metabolite diversity, we investigated *buxicola*, *decemcellulare* and *solani* species complexes in order to unveil missing enzymatic capabilities. We found that several genes in the said species complexes could contain enzymatic activities needed to explain the chemodiversity observed in the metabiosynthetic pathway.

Keywords: *Fusarium*, Polyketide synthase, metabiosynthetic pathway, gene cluster, evolution, *PGL1*, *fsr1*

5.2 Introduction

The ascomycete genus *Fusarium* includes both species with a saprophytic and plant pathogenic life style. The pathogenic members are responsible for significant agricultural losses for many important crops, via wilts and rot diseases and due to their production of mycotoxins that accumulates in the harvested crops [1, 2]. Their importance has fostered a whelm of studies into the individual species mode of infection, production of small molecules and the toxicology of these molecules. The investigated secondary metabolism within fusaria includes prominent mycotoxins includes fumonisins, zearalenone and trichothecenes [3], as well as many different pigments such as aurofusarin, bikaverin and fusarubins [4, 5, 6]. In the last decade focus has been on elucidating the genetic basis and understanding of the biosynthetic pathways responsible for production of these secondary metabolites, fuelled by the ever increasing availability of whole genome sequences for fusaria species. Comparative genomics studies focused on the distribution of polyketide synthase (PKSs) has revealed a huge diversity within the genus, and showed that only a three of the 67 different PKSs are conserved among all members of the genus [7]. The unusually conserved PKS gene, *pks3/pgl1/fsr1*, has via several studies been found to be responsible for formation of the dark perithecial pigmentation of many *Fusarium* spp. (all members of the former *Giberella* genus) [8] and for mycelium pigmentation in other species[4]. The chemical nature of the experimentally proven PGL1-produced pigments includes the fusarubin [4] and bostrycoidin compound families [9].

Several biosynthetic models for the formation of fusarubins, bostrycoins and structurally related compounds in *Fusarium* spp. have previously been proposed. The first model, proposed by Gatenbeck and Bentley (1965), suggest that javanicins and fusarubins are formed from an aromatic acid, which is converted into fusarubinoic acid and javanicin [10]. The second biosynthetic model was conceived by Arsenault (1968) and goes into a description of the biosynthesis of solaniol and bostrycoidin as well as fusarubin and javanicin [11]. The third proposed biosynthesis was formulated by Kurobane *et al.* (1980), in which a common intermediate of fusarubinalcohol leads to the formation of dihydrofusarubin, which is then converted to bostrycoidin, fusarubin, norjavanicin, and anhydrofusarubin [12]. Recently, two models have been made on the basis of genetic studies of *F. fujkuroi* and *F. graminearum*. Studt et al. (2012) investigated biosynthesis of fusarubins and related compounds in *F. fujikuroi* and found 6-O-demethylfusarubinaldehyde as

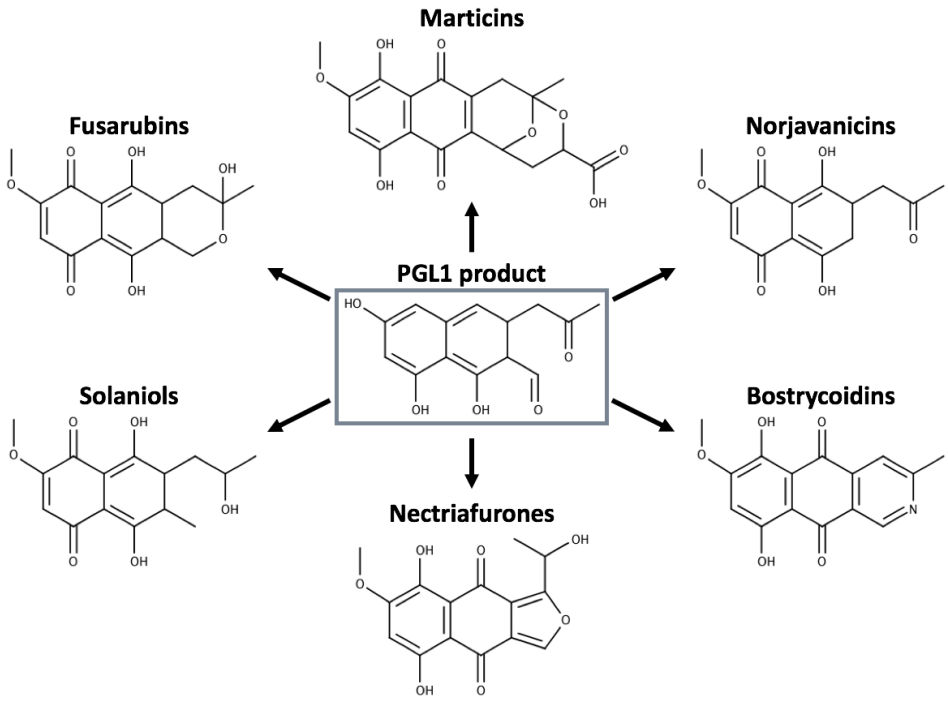


Figure 5.1 Chemical diversity originating from the *PGL1* gene cluster.

the first stable intermediate and likely the product of the PKS PGL1, named FSR1 in *F. fujikuroi* [4]. Frandsen *et al.* (2016) investigated biosynthesis of perithecial pigmentation of *F. graminearum* and found 6-O-demethyl-5-deoxybostrycoidin anthrone to be the first stable intermediate [9]. The naphthoquinone structure makes many of the *Fusarium* naphthoquinones antibacterial, but some metabolites have also been found to be phytotoxic and fungicidal [13, 14].

The fusarubin and bostrycoidin metabolite families share carbon core structures with that of solaniols, nectriafurones, marticinins and javanicins, which have been isolated from several other members of the *Fusarium* genus. Some of the chemical diversity of the PGL1 gene cluster is shown in figure 5.1. All members of these compound families include a naphthoquinone scaffold, but at the same time features distinctly different molecular features in their structure. The pattern of decoration of the core structures across the different metabolite families is in addition surprisingly similar [13]. With these structure similarities we hypothesized that all of these compounds have the same biosynthetic origin, but are the results of different evolutionary trajectories in the different subgroups of the genus.

Therefore, we wanted to investigate whether the chemical and underlying biosynthetic diversity within the *Fusarium* genus was reflected in the structure and enzymatic potential encoded by the *pgl1* gene cluster across the genus. Secondly, we set out to formulate a common biosynthetic model (metabiosynthetic pathway) for the *pgl1* cluster derived compounds with the aim of describing the theoretical minimum number enzymatic reactions that can explain the chemical diversity found across the whole genus of *Fusarium*.

5.3 Materials and methods

5.3.1 Bioinformatic analysis of *pgl1* gene cluster

The 343 analyzed *Fusarium* genome sequences (see appendix for full list) included all the *Fusarium* genomes, publicly available via NCBI, JGI, Broad/MIPS, along with recently sequenced isolates. Sequencing was done by following methods described in Brown and Proctor (2016) [7]. In breif, DNA extraction was done by using ZR Fungal/Bacterial DNA MiniPrepTM Kit (Zymo Research, Irvine, California) and a NExtera XT DNA Library Preparation Kit (Illumina, San Diego, California) to prepare the DNA libraries for the Iillumina MiSeq. The Genomic data was trimmed and the genomes were *de novo* assembled using CLC Genomics Workbench (CLC bio, Qiagen, Aarhus, Denmark). Whole genome gene predictions were performed using AUGUSTUS (http://augustus.gobics.de/).

The genomes were first analyzed by antiSMASH with the aim of identifying the *pgl1* gene cluster in the individual species [15]. Functional annotation for the encoded enzymes in the different *pgl1* gene clusters were made using Blast2GO [16]. A list of all analyzed strains, their source and antiSMASH analysis output can be found in appendix (Table S5.1). A subset of 22 isolate (Appendix Table S5.2 for strain table) were selected for further investigation, based on representing the whole genus of *Fusarium* and difference in gene cluster topologies. The existing automatically generated gene calls (AUGUSTUS) were manually inspected via multiple sequence alignments, using CLC Genomics Workbench (CLC bio, Qiagen, Aarhus, Denmark), of the nucleotide sequences with that of gene models for the same gene in the whole subset. This was done to ensure the same start and stop codon was used for all genes. When discrepancies in gene calls were found the gene call having highest consensus in the subset was also made for the sequences with deviating gene calls. Next, a nucleotide sequence synteny analysis of the *pgl1* gene cluster was made for the selected subset using Easyfig [17], with BLAST

parameters set to a bit size of minimum 50 bp and a minimum identity of 50%. To search for enzymatic activities, functional annotation of the encoded enzymes, which were not part of the *pgl1* clusters, was made using Pfam domain search [18] in CLC Genomics Workbench (CLC bio, Qiagen, Aarhus, Denmark).

5.4 Results and discussion

The pigmentation of fusaria has been investigated thoroughly and three polyketide compound groups have been found to be used as mycelial pigmentation for different subclades of the *Fusarium* genus. Members of the *Nectria* clade produces yellow and red mycelial pigmentation which have a naphthoquinone structure and is produced by the *pgl1* gene cluster [19]. Species in the *fujikuroi* and *oxysporum* species complexes produce the red bikaverin as the main mycelial pigmentat via the *bik1* gene cluster [5], and the members of the *sambucinum* and *tricinctum* species complexes owe their red mycelial color to aurofusarin produced by the gene cluster of *pks12* [6]. All species in the *Gibberella* clade have dark perithecia and get this pigmentation via the action of the *pgl1* gene cluster [8]. The produced pigment of perithecia has been proposed to be the dark purple purpurfusarin [9]. While the chemical composition of perithecial pigmentation for the *Nectria* clade is still unknown, deletion of the gene *pksN* produces colorless perithecia [20]. A thorough overview of naphthoquinones produced by fusaria and other fungi can be found in the review by Medentsev and Akimenko (1998) [13].

5.4.1 The *pgl1* gene cluster

In previous genetic studies of the genus *Fusarium* the *pgl1* backbone gene has been found for all investigated species [7, 21, 22]. For the investigated sequences in this study a different pattern was observed. To search for the *pgl1* gene cluster in the available genome sequences they were analyzed using antiSMASH. Based on the species complexes, as defined by O'Donnell *et al.* (2013) [21] four of the 20 species complexes were found not to have the *pgl1* backbone gene, namely *staphyleae*, *albidum*, *ventricosum* and *dimerum*, while all other species complexes did (See appendix Figure S5.1 for phylogeny of all isolates). The most parsimony explanation for this distribution is that the gene cluster was gained as the *solani* species complex evolved and then subsequently lost when the *staphyleae* species complex branched off from the rest. Previous studies have investigated the *pgl1* gene cluster and delimited it to consist of six genes *pgl1* to *pgl6*. Transcriptional

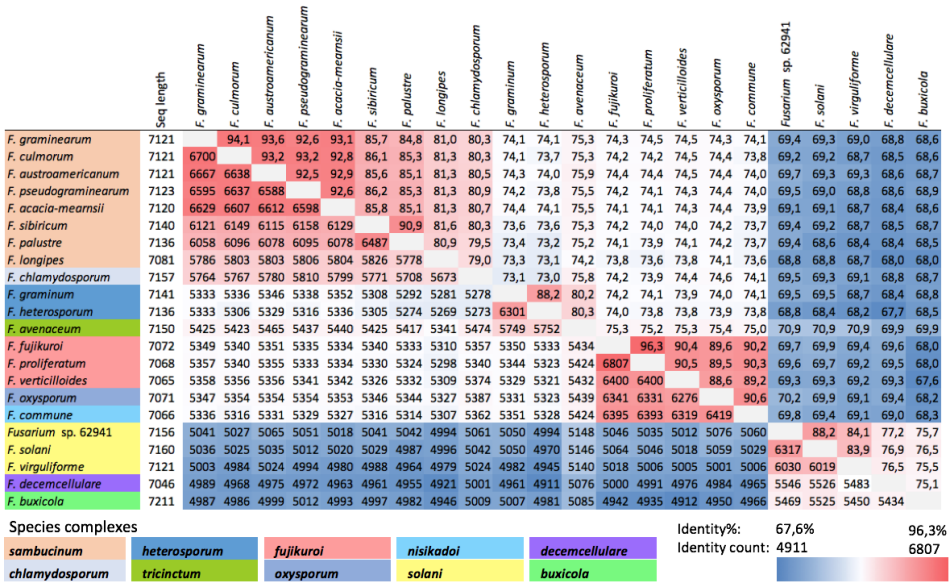
analysis in *F. graminearum* showed that six genes were co-regulated by the PGL6 transcription factor, named PglR in *Fusarium graminearum* [9], and transcriptional analysis of *F. fujikuroi*, grown under *pgl1* cluster inducing conditions also showed co-transcription of the same six genes [4]. An overview of the genes in the cluster can be found in table 5.1.

A subset (22 isolates) of the genomes investigated in the initial dataset were selected for further analysis of the *pgl1* gene cluster across the *Fusarium* genus. The subset of genomes was selected firstly based on the availability of genomic scaffolds that included the whole *pgl1* gene cluster. Secondly, to reflect the variability in cluster topology and finally, to include representatives for the different species complexes within the *Fusarium* genus (See appendix table S5.1 for antiSMASH output). Pairwise nucleotide sequence similarity analysis of the *plg1* cluster genes were done (Figure 5.2 (*plg1*) and appendix (*pgl2* to *pgl6*)). The highest similarity within genes was found between species of the same or closely related species complexes as defined by O'Donnell *et al.* (2013) [21].

Pairwise analysis of the nucleotide sequence identity for all genes in the various *pgl1* gene cluster (Appendix figure S5.2 to S5.6). Revealed the existence of four distinct groups, which followed the species complexes defined by O'Donnell *et al.* (2013) [21]. The first group included clusters from members of the *sambucinum* and *chlamydosporum* species complexes, while *tricinctum* and *heterosporum* formed a second group group. Then *fujikuroi*, *niskadoi* and *oxysporum* species complexes the third, and finally *buxicola*, *decemcellulare* and *solani* species complexes formed a fourth group. The *pgl1* gene of the fourth group was very different from three the other groups. This is in agreement with the fourth group being of the *Nectria* clade in the *Fusarium* genus and first, second and third group belonging to the *Gibberella* clade. To further investigate the *pgl1* gene cluster, a synteny

Table 5.1 The genes in the *pgl1* cluster in the investigated *Fusarium* species. Length bp is the maximum and minimum length in the subset of the investigated species. The enzymatic domains for the biosynthetic genes in *F. fujikuroi* IMI58289 is given [4].

Gene name	Length of gene (bp)	Conserved comains in the resulting enzyme (*F. fujikuroi*)	Predicted function of enzyme
pgl1	7046-7211	SAT, KS, AT, PT, ACP, ACP, R	PKS
pgl2	1206-1256	SAM-dependant methyl transferase	O-methyl transferase
pgl3	1640-1722	FAD-dependent oxidoreductase	Monooxygenase
pgl4	1196-1243	NAD(P)H dependent enoyl-reductase like	Oxidoreductase
pgl5	795-807	Short-chain dehydrogenase	Oxidoreductase
pgl6	990-1188	$Zn(II)_2Cys_6$ transcription factor	Transcription factor



	Seq length	F. graminearum	F. culmorum	F. austroamericanum	F. pseudograminearum	F. acacia-mearnsii	F. sibiricum	F. palustre	F. longipes	F. chlamydosporum	F. graminum	F. heterosporum	F. avenaceum	F. fujikuroi	F. proliferatum	F. verticilloides	F. oxysporum	F. commune	Fusarium sp. 62941	F. solani	F. virguliforme	F. decemcellulare	F. buxicola
F. graminearum	7121		94,1	93,6	92,6	93,1	85,7	84,8	81,0	80,3	74,1	74,1	75,3	74,3	74,5	74,5	74,3	74,1	69,4	69,3	69,0	68,8	68,6
F. culmorum	7121	6700		93,2	93,2	92,8	86,1	85,3	81,3	80,3	74,1	73,7	75,3	74,2	74,2	74,5	74,4	73,8	69,2	69,2	68,7	68,5	68,6
F. austroamericanum	7121	6667	6638		92,5	92,9	85,6	85,1	81,3	80,5	74,3	74,0	75,9	74,4	74,4	74,5	74,4	74,0	69,7	69,3	69,3	68,6	68,7
F. pseudograminearum	7123	6595	6637	6588		92,6	86,2	85,3	81,3	80,9	74,2	73,8	75,5	74,2	74,1	74,3	74,4	74,0	69,5	69,0	68,8	68,6	68,9
F. acacia-mearnsii	7120	6629	6607	6612	6598		85,8	85,1	81,3	80,7	74,4	74,1	75,5	74,1	74,1	74,3	74,4	73,9	69,1	69,1	68,7	68,4	68,6
F. sibiricum	7140	6121	6149	6115	6158	6129		90,9	81,6	80,3	73,6	73,6	75,3	74,2	74,0	74,0	74,2	73,7	69,4	69,2	68,7	68,5	68,7
F. palustre	7136	6058	6096	6078	6095	6078	6487		80,9	79,5	73,4	73,2	75,2	74,1	73,9	74,1	74,2	73,7	69,4	68,6	68,4	68,4	68,5
F. longipes	7081	5786	5803	5803	5806	5804	5826	5778		79,0	73,3	73,1	74,2	73,8	73,6	73,8	74,1	73,6	68,8	68,8	68,7	68,0	68,0
F. chlamydosporum	7157	5764	5767	5780	5810	5799	5771	5708	5673		73,1	73,0	75,8	74,2	73,9	74,4	74,6	74,1	69,5	69,3	69,1	68,8	68,7
F. graminum	7141	5333	5336	5346	5338	5352	5308	5292	5281	5278		88,2	80,2	74,2	74,1	73,9	74,0	74,1	69,5	69,5	68,7	68,4	68,8
F. heterosporum	7136	5333	5306	5329	5316	5336	5305	5274	5269	5273	6301		80,3	74,0	73,8	73,8	73,9	73,8	68,8	68,4	68,2	67,7	68,5
F. avenaceum	7150	5425	5423	5465	5437	5440	5425	5417	5341	5474	5749	5752		75,3	75,2	75,3	75,4	75,0	70,9	70,9	70,9	69,9	69,9
F. fujikuroi	7072	5349	5340	5351	5335	5334	5340	5333	5310	5357	5350	5333	5434		96,3	90,4	89,6	90,2	69,7	69,9	69,4	69,6	68,0
F. proliferatum	7068	5357	5340	5355	5333	5334	5330	5324	5298	5340	5344	5323	5424	6807		90,5	89,5	90,3	69,6	69,7	69,2	69,5	68,0
F. verticilloides	7065	5358	5356	5356	5341	5342	5326	5332	5309	5374	5329	5321	5432	6400	6400		88,6	89,2	69,3	69,3	69,2	69,3	67,6
F. oxysporum	7071	5347	5354	5354	5354	5353	5346	5344	5327	5387	5331	5323	5439	6341	6331	6276		90,6	70,2	69,9	69,1	69,4	68,2
F. commune	7066	5336	5316	5331	5329	5327	5316	5314	5307	5362	5351	5328	5424	6395	6393	6319	6419		69,8	69,4	69,1	69,0	68,3
Fusarium sp. 62941	7156	5041	5027	5065	5051	5018	5041	5042	4994	5061	5050	4994	5148	5046	5035	5012	5076	5060		88,2	84,1	77,2	75,7
F. solani	7160	5036	5025	5035	5012	5020	5029	4987	4996	5042	5050	4970	5146	5064	5046	5018	5059	5029	6317		83,9	76,9	76,5
F. virguliforme	7121	5003	4984	5024	4994	4980	4988	4964	4979	5024	4982	4945	5140	5018	5006	5005	5001	5006	6030	6019		76,5	75,5
F. decemcellulare	7046	4989	4968	4975	4972	4963	4961	4955	4921	5001	4961	4911	5076	5000	4991	4976	4984	4965	5546	5526	5483		75,1
F. buxicola	7211	4987	4986	4999	5012	4993	4997	4982	4946	5009	5007	4981	5085	4942	4935	4912	4950	4966	5469	5525	5450	5434	

Figure 5.2 Identity plot showing the nucleotide identity of the *pgl1* genes from different species within the *Fusarium* genus. The lower triangle shows the number of identities while the upper triangle show the identity percentage. The species complexes as well as species complex groups are highlighted.

plot of the cluster was made for the subset of genome sequences (Figure 5.3).

The synteny analysis of genes in the *pgl1* gene cluster showed that seven genes were highly conserved across the *Fusarium* genus, which was supported by the identity plots of the biosynthetic genes (see figure 5.2 and appendix). The sequence and direction of the genes from *pgl1* to *pgl6* was identical for all analyzed sequences except for two sequences. In *F. longipes* *pgl3* had translocated downstream of *pgl1* and in *Fusarium* sp. 62941 two parts of the cluster, *pgl1* to *pgl3* and *pgl4* to *pgl6* was inverted compared to other sequences. Within the *sambucinum* species complex the sequence synteny plot showed a very high degree of sequence conservation of the whole gene cluster including both coding and non-coding regions. Atypically, this level of conservation was also observed to transcend the species complex boundaries for between *fujikuroi*, *nisikadoi* and *oxysporum* species complexes. The topology of the cluster in members of the solani and decemcellulare species complexes differed significantly from that of the other isolates, as the cluster was split in two and the two parts were separated by 12-17 genes between pgl3 and pgl4. This was also previously noted by Frandsen et al. (2016) for *F. virguliforme* and *F. solani* [9]. The genes separating the two halves of the cluster

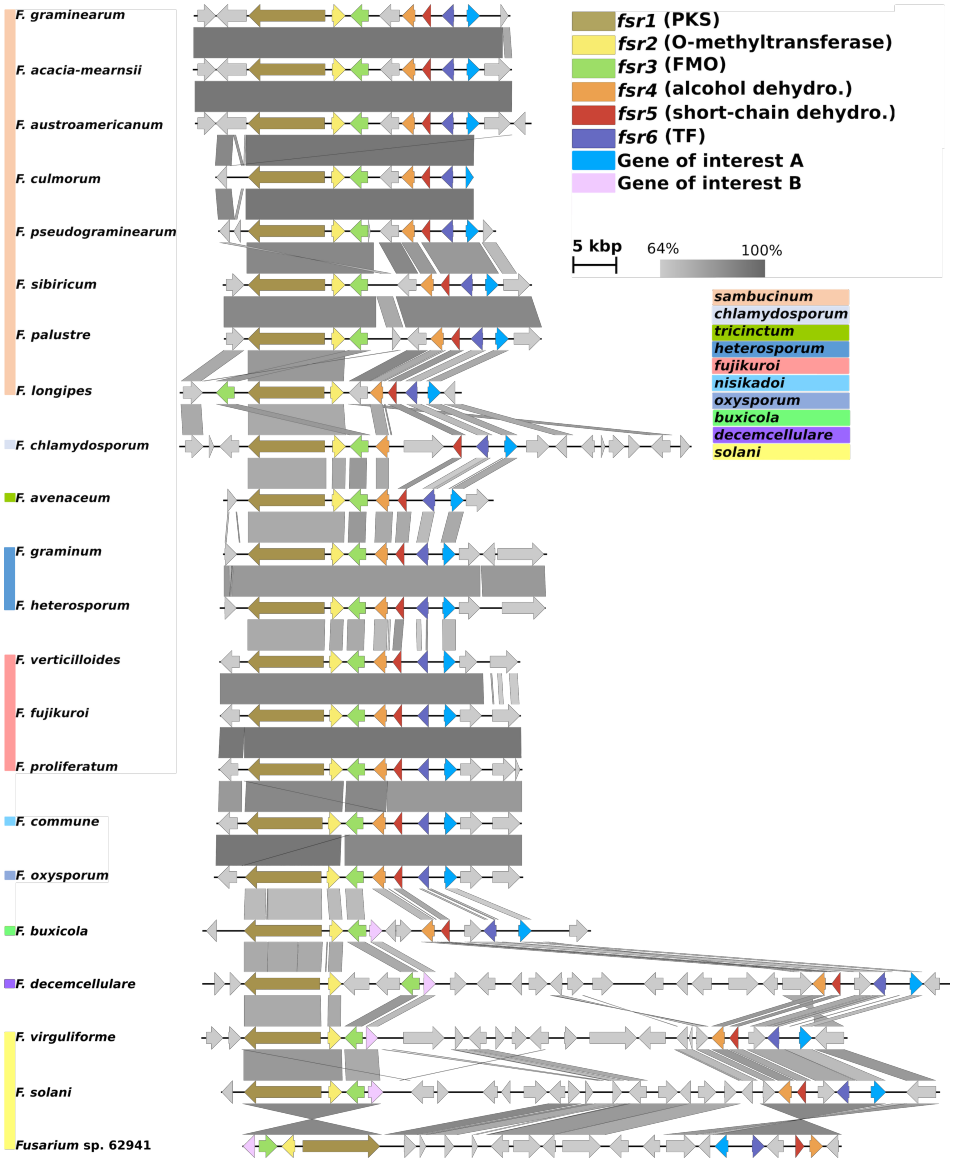


Figure 5.3 Synteny plot of the *pgl* gene cluster across the *Fusarium* genus. The color coding alongside the species names refer to the species complexes in figure 5.2. Genes which are not part of the *pgl1* biosynthetic cluster are grey.

only shows limited synteny between the species. For *Fusarium* sp. 62941 the gene synteny with *F. solani* was quite high for the *pgl1* cluster and the genes between *pgl3* and *pgl4*, but the two parts of the cluster was inverted compared to *F. solani*.

The same pattern of clustering of the biosynthetic genes in all species complexes except the *solani* species complex was in good agreement with the findings in Frandsen *et al.* (2016) [9]. Also, we found that *decemcellulare* and to some extent *buxicola* species complexes also had genes inserted between the two parts of the cluster. As the *buxicola* species complex is closer related to *oxysporum* and *fujikuroi* species complexes not having the aforementioned insert, it could be hypothesized that *F. buxicola* represents an intermediate stage of the *pgl1* cluster development towards becoming clustered. The gene downstream of *pgl6* in *F. graminearum*, which was named *pglE* in Frandsen et al. (2016) and named Gene of interest A here, was found not to be co-regulated with the other genes in the cluster in *F. graminearum* [9] nor in *F. fujikuroi* [4]. However, the conservation of the gene in all the investigated clusters suggests that the encoded enzyme could be important for the function of the cluster as other genes around the cluster genes change throughout the *Fusarium* genus. The presence of an esterase domain in *pglE* could indicate that the gene has a function in the biosynthetic pathway, but that it is regulated independently of the other genes in the cluster. The majority of compounds that have been shown to or hypothesized to be products of the *pgl1* cluster have been isolated from *Fusarium* species belonging to the *solani* species complex. Interestingly members of the *solani*, *buxicola* and *decemcellulare* species complex share one addition conserved gene, named Gene of interest B, located downstream of *pgl3* in these species. This gene encodes an enzyme with an alcohol dehydrogenase domain, an activity which is often observed in the tailoring of aromatic polyketides. The conservation of the gene and the relevant enzymatic potential makes the gene a good candidate for being involved in the biosynthesis of PGL1 derived compounds in these species, and possibly explaining part of their unique chemical diversity.

5.4.2 The metabiosynthetic pathway of PGL1-derived compounds

Many of the metabolites associated with the *pgl1* gene cluster has been observed as extracellular [4, 24] or perithecial pigments [8, 9], while others have also been found to be toxins coupled to pathogenicity in plants on which fusaria are often

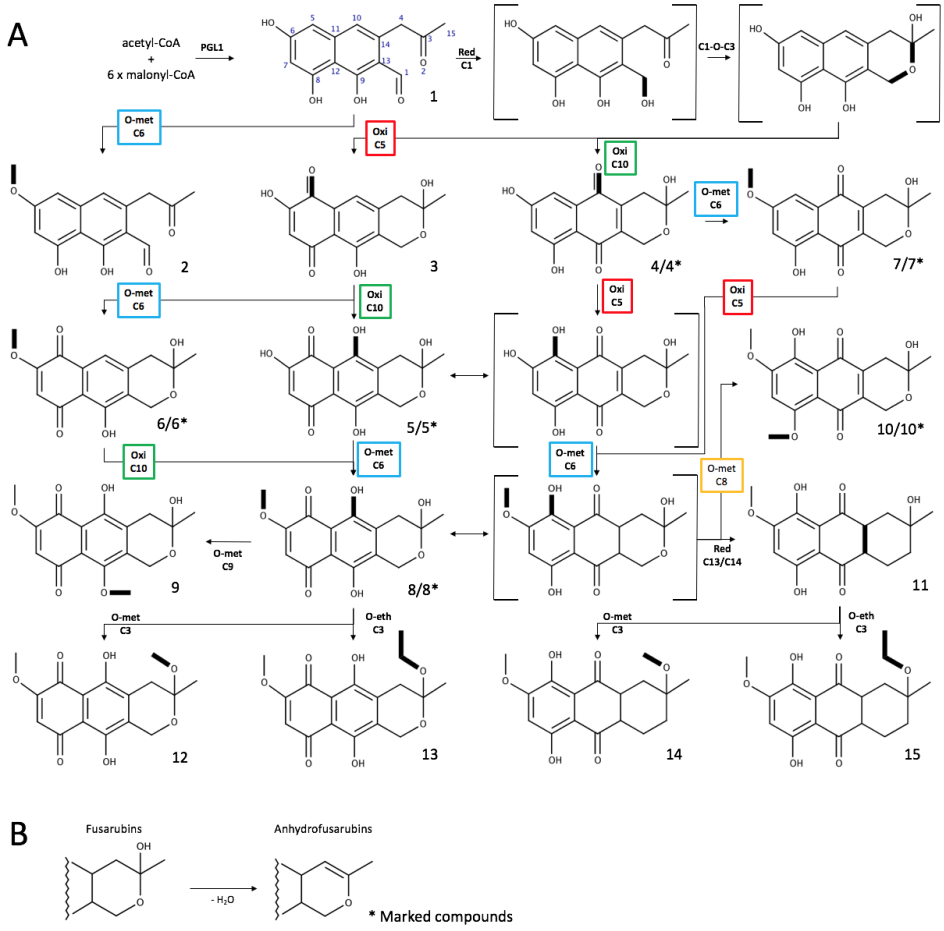


Figure 5.4 Proposed metabolic pathway leading to chemodiversity of fusarubins and anhydrofusarubins in *Fusarium* spp. All compounds are described in Medentsev and Akimentko (1998), Studt *et al.* (2012) and Parisot *et al.* (1991) [4, 13, 23] A. Proposed biosynthesis of fusarubins. Compounds with asterics (*) have also been isolated as anhydrofusarubins. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9. B. Conversion between fusarubins and anhydrofusarubins by C3 reduction.

found [25, 26]. To create a more detailed biosynthetic model and to identify general biosynthetic conversions, we searched the available literature, and identified bostrycoidin- and fusarubin-like compounds produced by *Fusarium* spp. Based on their shared substructural features, we formulated a metabiosynthetic model including 47 known compounds and a number of theoretical intermediates. In the following proposed biosynthetic pathways the carbon numbering system follows the system shown in Studt *et al.* (2012) [4].

The initial product from the PKS PGL1 has been found to be 6-O-demethylfusarubinaldehyde (compound **1**), based on heterologously expression of the *F. solani pgl1* in *Aspergillus oryzae* [19]. Furthermore, deletion of the flavin-dependant monooxygenase (FMO) encoding gene *pgl3* in *F. fujikuroi* resulted in the accumulation of only fusarubinaldehyde (compound **2**) [4], which showed that the oxidation on C5 and C10 is required for biosynthesis of fusarubins and anhydrofusarubins (Figure 5.4). While oxidation seems to be the required for production of later stage compounds, deleting the O-methyltransferase (O-MT) *pgl2* in *F. fujikuroi* yielded the C5 oxidized 6-O-demethyl-10-deoxyfusarubin (compound **3**) and 6-O-demethylfusarubinaldehyde (compound **2**) [4]. Here it can be concluded that the fusarubin ring formation can happen without the C6 O-methylation. The opposite biosynthetic sequence of reactions was observed for formation of bostrycoidins in *F. graminearum*, where the 6-O-demethyl-5-deoxybostrycoidin anthrone (compound **16**) was observed when overexpressing the PKS *pgl1* [9]. This compound has the bostrycoidin ring confirmation while it does not have C6 methylation nor C5 or C10 oxidation. An overview of our proposed biosynthesis in the *Fusarium* genus can be found in figure 5.4 for fusarubins and anhydrofusarubins, and in figure 5.5 for bostrycoidins.

For formation of the pyran ring in fusarubins the C1 carbon needs to be reduced to an alcohol, to allow formation of the pyran ring via acetal formation. This C1 reduction is hypothesized to be carried out by the FMO *pgl3* by Studt *et al.* (2012) [4]. However, we would find it surprising and unlikely that the same FMO would do both oxidations and reductions on the same molecule and on vastly different parts of the molecule. Hence, we expect that one or more enzymes are responsible for the C1 reductions, but the identity of the responsible enzyme or enzymes remains unknown. We hypothesize that the conversion between fusarubins and anhydrofusarubins to be driven by an enzyme to create the C3/C4 double bond in anhydrofusarubins. An alternative explanation of this

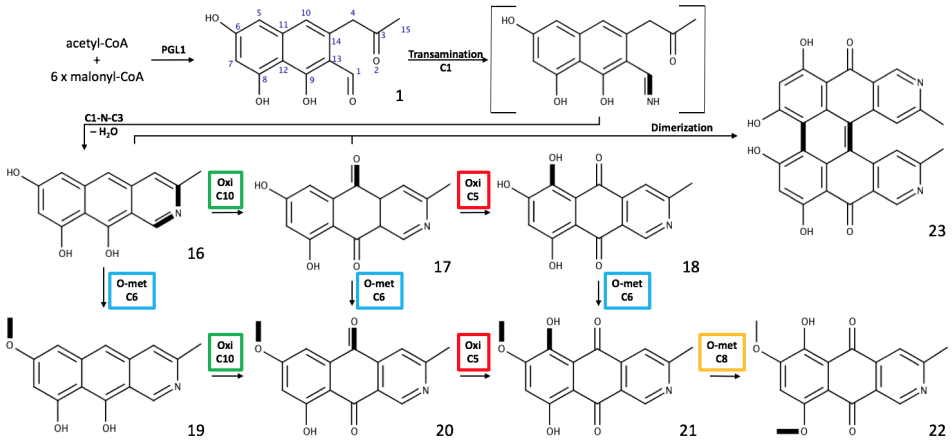


Figure 5.5 Proposed biosynthesis of bostrycoidins produced by *Fusarium* spp. All compounds are described in Frandsen *et al.* (2016) and Medentsev and Akimentko (1998) [9, 13]. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9.

double bond could be formed by high acid concentrations during purifications as culture filtrate fractions often were acidified prior to extraction [27, 28, 29]. A step in the biosynthesis of bostrycoidins are also missing further investigation. Parisot *et al.* (1989) observed that incorporation of nitrogen in bostrycoidins can happen spontaneously by co-incubation of ammonia and fusarubin lactol (compound **37**), even though free ammonia is unlikely to be present inside the cell [30]. A similar formation of a pyridine ring is observed in monascus pigments by Chen *et al.* (2017) [31], where incorporation of nitrogen can also happen via incubation of a precursor containing a pyran moiety with ammonia. For monascus pigments the nitrogen does not have to originate from ammonia, but can also come from free amino acids [31]. An alternative bostryocoidin biosynthetic route was proposed by Wagoner *et al.* (2008) [32], which is also shown in our proposed biosynthesis of bostrycoidins (Figure 5.5). In this route the conversion is initiated with enzymatic transamination of the C1 aldehyde to an imine, which then via nucleophilic attack on the C3 carbonyl group followed by water loss forms the pyridine ring.

Dihydrofusarubins (compounds **11**, **14** and **15)** which have the reduced C13/C14 bond, have only been isolated from *F. solani* [14, 28]. Dihydrofusarubin has been hypothesized to be a precursor of fusarubin by Kurobane *et al.* (1980) as it spontaneously was converted to fusarubin when exposed to oxygen under basic conditions [12]. An alternative hypothesis could be that dihydrofusarubin is formed by re-

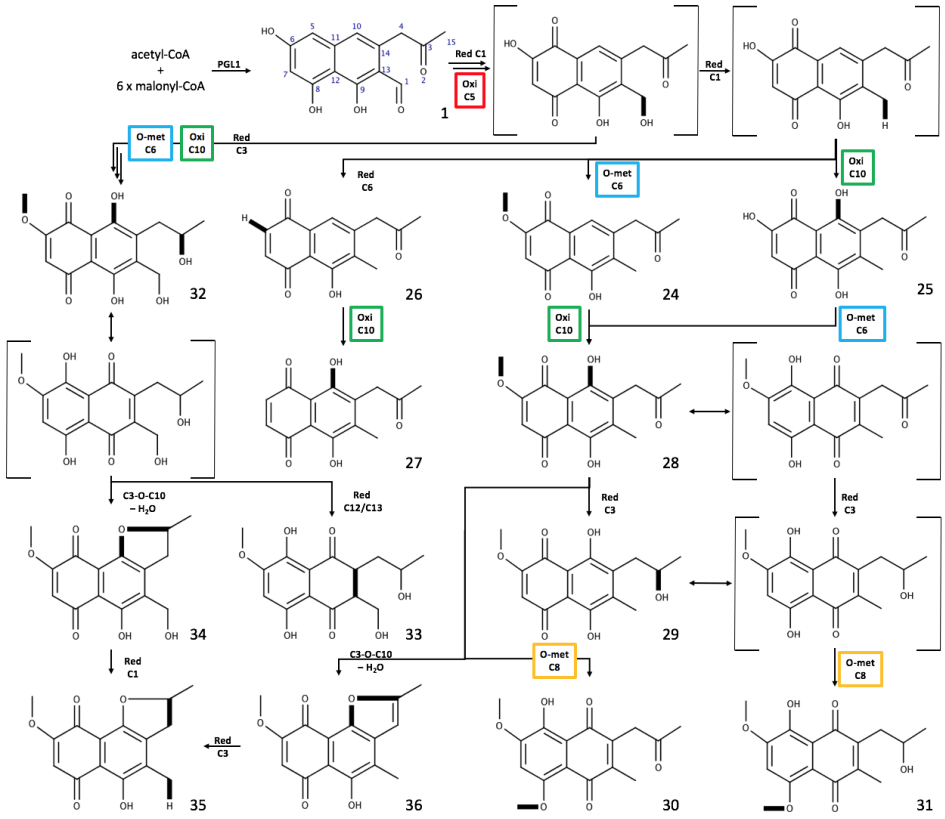


Figure 5.6 Javanicins and solaniols produced by *Fusarium* spp. All compounds are described in medentsev and Akimentko (1998) and Parisot *et al.* (1990) [13, 33]. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9.

duction of the C13/C14 bond after formation of fusarubin. As dihydrofusarubin only has been isolated from *F. solani* the gene responsible for the C13/C14 reduction may be unique to *F. solani*. The existence of C3 methylation and ethylation of fusarubin (compounds **12** and **13**) and dihydrofusarubin (compounds **14** and **15**) has been hypothesized to be artifacts, which would occur during chemical purification. Nonetheless, Kurobane *et al.* (1986) obtained these compounds while purification was done without the presence of methanol and ethanol which could lead to formation of methyl- and ethyl ethers [14]. The absence of a characterized ethyl transferase in nature leaves some speculation on the biosynthesis of the C3 ethyl compounds. The formation of the ethyl may happen via first O-methylation and then addition of another carbon by a different C-methyl transferase.

Another group of metabolites derived from the *pgl1* gene cluster are the solaniols and javanicins. For production of solaniols and javanicins the C1 aldehyde would have to be reduced to an alcohol or to an methyl group (Figure 5.6). The sequence of reduction of C1 and the C5 and C10 oxidation and C6 O-methylation appears to be in favor of the latter being the initial one as most solaniols and javanicins have been isolated as having both of the C5 and C10 oxidations and the C6 O-methylation. An exception to this observation is 10-deoxyjavanicin (compound **24**). Interestingly this metabolite along with 5- or 10-deoxyfusarubin (compound **6** or **7**) and 5- or 10-deoxyanhydrofusarubin (compound **6*** or **7***) have been isolated from yellow mutants of *F. solani* [23, 34]. The accumulation of the 5- and 10-deoxy compounds in these yellow mutants could offer a quick avenue, via genome sequencing, to identify other genes which may modulate the activity of the FMO *pgl3*. Therefore, it could be interesting to investigate the genome of these yellow mutants and search for the mutations leading to the accumulation of the 5- and 10-deoxy compounds.

The biosynthesis of solaniols and javanicins requires the C1 aldehyde to be reduced to a methyl, and furthermore, the C3 carbonyl to be reduced to an alcohol. Also, the 6-deoxynaphthoquinones (compounds **26** and **27**) is dependent on another enzymatic activity, which is able to reduce the C6 alcohol. The anhydrojavanicins (compounds **34**, **35** and **36**) have only been isolated from *F. solani* or the closely related *F. decemcellulare* [13, 27]. As with dihydrofusarubins, this could mean that specific genes are needed for the alternative ring formation seen in anhydrojavanicin. These may have been lost or gained as the cluster evolved throughout the *Fusarium* genus to become more clustered as seen in the Gibberella clade. Formation of dihydroanhydrojavanicin and 1-hydroxydihydroanhydrojavanicin (compounds **34** and **35**) both depends on reduction of the C3 position and it would therefore make more sense to name these as anhydrosolaniols instead of anhydrojavanicins.

The next group of metabolites which will be put into the biosynthetic context of the *pgl1* gene cluster here, are produced by ringformation of the C1 aldehyde with the C3 ketone to form fusarubin lactol (compound **37**) or by oxidation of C1 aldehyde to form fusarubinoic acid (compound **39**), which can then undergo ring formation with the C3 ketone to make fusarubin lactone (compound **40**). Through decarboxylation, fusarubinoic acid can also yield norjavanicin (compound **41**), which can finally be transformed to 13-hydroxynorjavanicins (compound **42**,

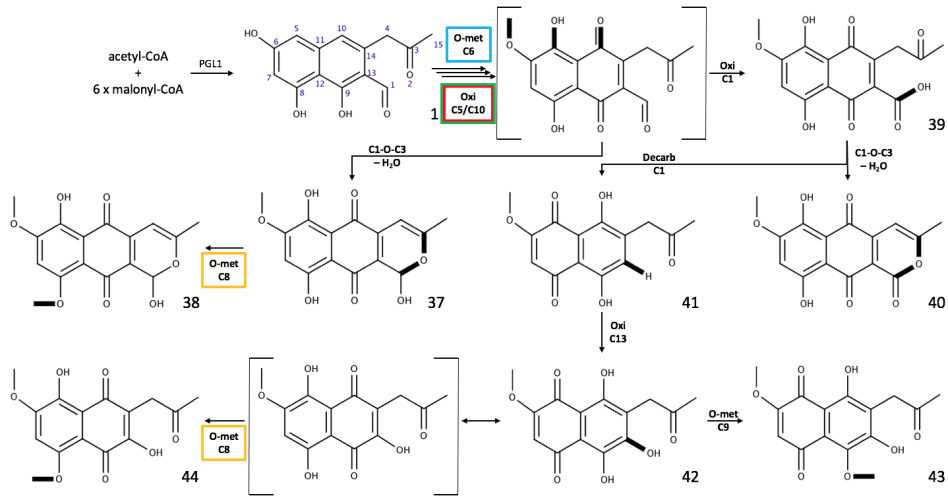


Figure 5.7 Fusarubin lactone, fusarubin lactol and norjavanicins produced by *Fusarium* spp. Compounds are described in Medentsev and Akimenko (1998) and Studt *et al.* (2012) [4, 13]. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9.

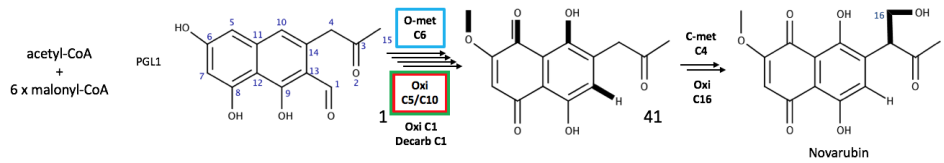


Figure 5.8 Proposed biosynthesis of novarubin by *Fusarium* spp. if derived from the *PGL1* pathway. Novarubin is described in Medentsev and Akimenko (1998) [13]. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9.

also named 6-hydronorjavanicins in literature) by C13 oxidation. The reductive release from PGL1 followed by C1 oxidation as seen for synthesis of fusarubinoic acid and later derivatized compounds is uncommon in biosynthesis of natural products. Nonetheless, fusarubinoic acid and the later derivatized compounds, ie. fusarubin lactone and norjavanicins which have been isolated in several different studies isolating naphtoquinone pigments from *Fusarium* spp. [13, 27, 35, 36].

A compound isolated from *Fusarium* spp., which share chemical features with the naphthoquinones described here, is novarubin. Previous reports of novarubin isolation has been from species in the *solani* and *decemcullulare* species complexes [13, 37]. Recently, novarubin has been detected, based on MS charateristics, in cultivations of wildtype *F. graminearum* [38], which has not been previously shown

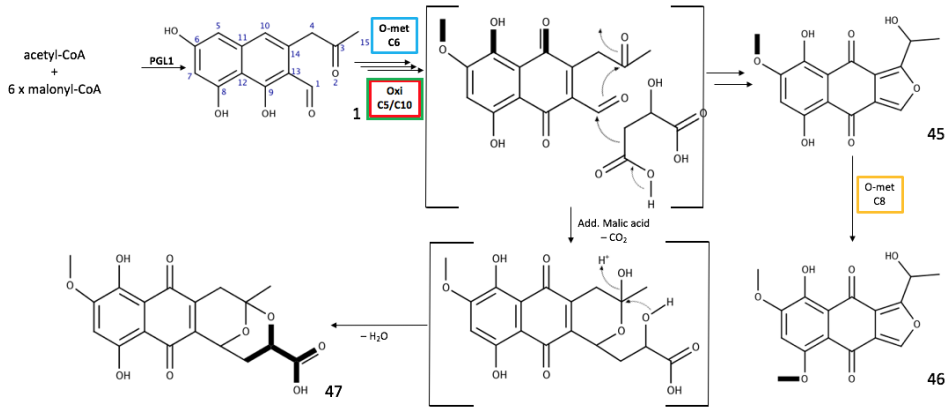


Figure 5.9 Proposed biosynthesis of nectrifurones and marticin produced by *Fusarium* spp. Compounds are described in Medentsev and Akimenko (1998), Studt *et al.* (2012) and Parisot *et al.* (1991) [4, 13, 23]. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9.

to produce fusarubins, javanicins or solaniols. This could lead one to speculate that the biosynthetic origin of novarubin is not from the *pgl1* cluster even though the basic structure is very similar to that of norjavanicin. If novarubin is derived from the pathway it is most likely via C4 methylation of norjavanicin (compound **41**) followed by oxidation of the same carbon (Figure 5.8).

The final group of metabolites for which a biosynthetic route will be proposed are nectriafurones and marticin (Figure 5.9). The biosynthesis of nectriafurones has been proposed by Studt *et al.* (2012) to be derived from an intermediate with the C1 in form of an alcohol, which then forms the furan ring in equilibrium [4]. It appears to be more likely that the precursor of nectriafurone is the C1 aldehyde as this would give the correct oxidation state prior to the furan ring formation as depicted in our biosynthesis, though a good mechanism for this formation is absent. The biosynthesis of marticin have been proposed to be addition of malic acid [39] and this is also depicted in our proposed biosynthesis.

In general compounds having the C8 or C9 O-methylations have rarely been isolated from *F. solani*, but instead they have been isolated from species in the *fujikuroi* and *oxysporum* species complexes. This pattern is observed for 8-O-methylnectriafurone (compound **46**), 8- and 9-O-methylfusarubin (compounds **9** and **10**), 8-O-methylanhydrofusarubin (compound **10***), 8- and 9-O-methyl-6-hydroxynorjavanicin (compounds **43** and **44**), 8-O-methylsolaniol (compound **31**),

8-O-methylbostrycoidin (compound **22**), and 8-O-methylanhydrofusarubinlactol (compound **38**) [4, 9, 13]. The only compound for which this is not observed are for 8-O-metyljavanicin (compound **30**) which has been isolated from *F. solani* [40] as well as *F. fujikuroi* and *F. oxysporum* [13]. This exception to the C8 O-methylation brings one to speculate that the identification of *F. solani* by Kimura *et al.* (1981) could be wrong and the isolate from which they isolated 8-O-metyljavanicin (compound **30**) could have been a *F. fujikuroi* or *F. oxysporum*. If true and the C8 and C9 O-methylations are indeed unique to *F. oxysporum* and *F. fujkuroi*, then the O-methyltransferase *pgl2* could be more promiscuous in *F. oxysporum* and *F. fujkuroi* as proposed by Studt *et al.* (2012) or a completely different enzyme or enzymes are responsible for the C8 and C9 O-methylations, which is only present in these species.

5.4.3 Comparison of the metabiosynthetic pathway and the *pgl1* biosynthetic gene cluster

When observing the biosynthetic pathways proposed in figure 5.4 to 5.9 a common biosynthetic pattern is observed. The C6 O-methylation and the C5 and C10 oxidations are found for the vast majority of the isolated compounds, as well as the C8 O-methylation as has also been highlighted in the figures. Therefore, the main differences between compounds are observed for the C1 reductions or oxidations as well as the different possible ring formations. In figure 5.10 we attempt to present the otherwise large and complicated metabiosynthetic pathway in as few chemical conversions as possible, by dividing the conversions into an A-side reactions for C1 to C4 and C15, and a B-side reactions for C5 to C14. Here focus has been put on the enzymatic conversions instead of the individual compounds. To link the chemodiversity presented the previous section with the genetic diversity investigated in section 5.4.1, we made an overview of the enzymatic conversions needed for the presented chemodiversity (Table 5.2).

Not all reactions listed in table 5.2 are necessarily enzyme catalyzed. An example is cyclisation reactions as some of these are likely to be spontaneous, but eight different cyclisation reactions are found in figure 5.10 and therefore some of them are likely to be enzyme directed. As an example, the formation of the nectriafurone furan ring are likely to be enzyme catalyzed as the furan ring formation puts more strain on the ring. Furthermore, the biosynthesis of marticin is likely to be enzyme catalysed, but an enzyme with such an activity is currently

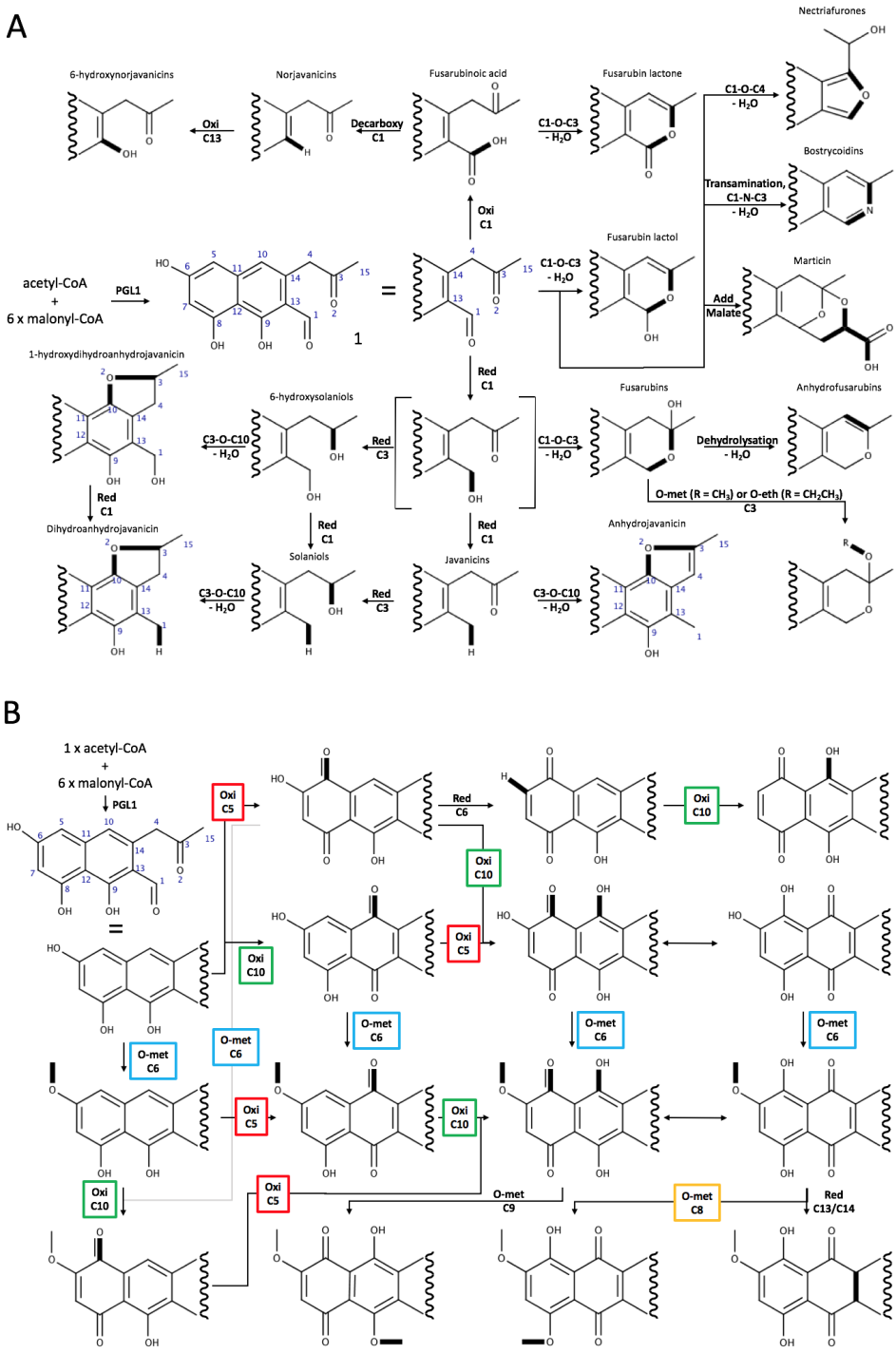


Figure 5.10 Simplified metabiosynthetic pathway representing the possible conversions of the A- and B-side of the PGL1 derived compounds. Emphasized bonds are the group on which the previous enzymatic reaction has happened. Boxed and coloured reactions are common for all described biosynthetic schemes in figures 5.4 to 5.9.

Table 5.2 Overview of enzymatic activities needed for putative metabiosynthetic pathway presented in figures 5.4 to 5.9.

Carbon #	Reaction	Putative/Known enzyme
1	Red CHO to CH_2OH	Reductase
1	Red CH_2OH to CH_3	Reductase
1	Oxi CHO to COOH	Oxidase
1	Decarboxylation	Decarboxylase
1	Transamination	Transaminase
1	Addition of malate	Unknown
1-3	Cyc C1-O-C3 or C1-N-C3	Unknown
1-4	Cyc C1-O-C4	Unknown
3	O-methylation or ethylation	O-MT
3	Red CO to CHOH	Reductase
3-10	Cyc C3-O-C10	Unknown
5	Oxi CH2 to CHOH	*pgl3* (FMO)
6	O-methylation	*pgl2* (O-MT)
6	Red CHOH to CH_3	Reductase
8	O-methylation	*pgl2*
9	O-methylation	*pgl2* or similar O-MT
10	Oxi CH_2 to CHOH	*pgl3* (FMO)
13	Oxi CH to COH	FMO
13-14	Red C=C to C-C	Reductase

unknown.

The number of presented biochemical conversions, which have been connected to a specific enzyme, is rather limited and leads one to speculate about how the observed chemical diversity is created. Therefore, the identity of at least five different enzymes for biochemical conversions of C1 is currently unknown being reduction, oxidation, decarboxylation, transamination and addition of malate. At least two biochemical conversions of C3 is currently not assigned to a specific enzyme being reduction and methylation or ethylation. Studt *et al.* also proposes that the O-methylase *pgl2* is responsible for O-methylations on C6 and C8 [4]. While O-methylation of C9 is not often observed this could also be catalysed by *pgl2* though we find this unlikely and therefore another O-methylase is needed for this conversion. Finally, two more enzymes are currently unknown. Namely the enzymes responsible for the C13 oxidation for biosynthesis of 6-hydronorjavanicins, which most like is catalyzed by an FMO and a reductase to reduce the C13-C14 double bond for dihydrodusarubin biosynthesis. This brings the total number of unknown enzymes to ten.

These unknown enzymatic activities could to some degree be explained by the biosynthetic enzymes being promiscuous and having low substrate and reaction specificity. This promiscuity can be used by an organism to diversify the metabolites and develop novel pathways [41]. An example of the promiscuity of the biosynthetic enzymes can be found for the C5 and C10 oxidation by *pgl3* and C6 and C8 O-methylation by *pgl2*. These are both, to a certain degree, inde-

pendent of what state other parts of the substrate is in. This can be concluded when regarding lack of sequence of C6 O-methylation and C5 and C10 oxidations observed for fusarubins (Figure 5.4) and bostrycoidins (Figure 5.5) and the oxidation and O-methylation appear to occur completely independently of one another. The only limitation appears to be in fusarubins where at least one of the C5 or C10 oxidations have to occur before C6 O-methylation. This difference can either be attributed to the difference of bostrycoidins and fusarubins or to difference in enzyme specificity in *F. fujikuroi* and *F. graminearum*.

An alternative explanation of unknown enzymatic activities could simply be that these genes have not been identified yet. The large distance between *pgl1* to *pgl3* and *pgl4* to *pgl6* in species complexes *buxicola*, *decemcellulare*, and *solani* could be due to more biosynthetic enzymes being part of the biosynthetic cluster in these species. To investigate this we did Pfam domain search in these coding sequences and the results for these can be found in appendix (Table S5.3). This search found several enzymes to have FMO or cytochrome P450 domains as well as several dehydrogenases and reductases. Several transcription factors were also identified, which could be controlling other parts of the biosynthetic genes for synthesis of PGL1 derived compounds, than what is controlled by Pgl6. Some methyltransferases were also identified. Finally, in *F. decemcellulare* an enzyme could be speculated to be linked to transamination, as the enzyme was found to be binding pyroxidal phosphate (PLP), which is known from transaminases [42]. However, this enzyme cannot be found in any of the other sequences investigated in this study, and a central role for biosynthesis of bostrycoidins is therefore unlikely.

Further studies are needed on the metabolism of the *pgl1* gene cluster in fusaria in order to fully understand the chemodiversity of the metabiosynthetic pathway presented here.

5.5 Conclusion and perspectives

This study set out to investigate the genetic diversity of the *pgl1* gene cluster and secondly to propose a metabiosynthetic pathway for compounds assumed to originate from the PGL1 biosynthetic pathway. We found that the *pgl1* biosynthetic cluster was highly conserved across the *Fusarium* genus with genes and topology being highly homologous for all species, except for species complexes *buxicola*, *decemcellulare* and *solani*. The metabiosynthetic put 47 previously isolated *Fusarium* spp. metabolites into one biosynthetic context and found common biosynthetic

steps in many parts of the pathway. The metabiosynthetic pathway also highlighted that the currently identified *pgl1* biosynthetic cluster would not suffice to explain the chemodiversity in the present pathway. Therefore, we searched in sequences from *buxicola*, *decemcellulare* and *solani* species complexes for genes possessing domains with relevant secondary metabolism activity and several genes were found to possess interesting biocatalysis domains. Transcriptome analysis of the *pgl1* gene cluster and neighbouring genes in the said species complexes could unveil enzymes with biosynthetic relevance for PGL1 derived compounds.

5.6 Acknowledgements

This work was funded by the Novo Nordisk foundation grant number NNF15OC0016626.

5.7 Author contributions

The research study was conceived by KK and RF. The bioinformatic analyses were designed and conducted by KK, HK and RP. Drafting of the manuscript was done by KK, and refined by RF and TL.

References

[1] Gordon, TR. "*Fusarium oxysporum* and the *Fusarium* wilt syndrome". In: *Annu Rev Phytopathol* 55 (2017), pp. 23–39.

[2] Ploetz, RC. "*Fusarium* wilt of banana". In: *Phytopathol* 105.12 (2015), pp. 1512–1521.

[3] Antonissen, G, Martel, A, Pasmans, F, Ducatelle, R, Verbrugghe, E, Vandenbroucke, V, Li, S, Haesebrouck, F, van Immerseel, F, and Croubels, S. "The impact of *Fusarium* mycotoxins on human and animal host susceptibility to infectious diseases". In: *Toxins* 6.2 (2014), pp. 430–452.

[4] Studt, L, Wiemann, P, Kleigrewe, K, Humpf, HU, and Tudzynski, B. "Biosynthesis of fusarubins accounts for pigmentation of Fusarium fujikuroi perithecia". In: *Appl Environ Microbiol* 78.12 (2012), pp. 4468–4480.

[5] Wiemann, P, Willmann, A, Straeten, M, Kleigrewe, K, Beyer, M, Humpf, HU, and Tudzynski, B. "Biosynthesis of the red pigment bikaverin in *Fusarium fujikuroi*: genes, their function and regulation". In: *Mol Microbiol* 72.4 (2009), pp. 931–946.

[6] Frandsen, RJN, Nielsen, NJ, Maolanon, N, Sørensen, JC, Olsson, S, Nielsen, J, and Giese, H. "The biosynthetic pathway for aurofusarin in *Fusarium graminearum* reveals a close link between the naphthoquinones and naphthopyrones". In: *Mol Microbiol* 61.4 (2006), pp. 1069–1080.

[7] Brown, DW and Proctor, RH. "Insights into natural products biosynthesis from analysis of 490 polyketide synthases from Fusarium". In: *Fungal Genet Biol* 89 (2016), pp. 37–51.

[8] Proctor, RH, Butchko, RAE, Brown, DW, and Moretti, A. "Functional characterization, sequence comparisons and distribution of a polyketide synthase gene required for perithecial pigmentation in some *Fusarium* species". In: *Food Addit Contam* 24.10 (2007), pp. 1076–1087.

[9] Frandsen, RJN, Rasmussen, SA, Knudsen, PB, Uhlig, S, Petersen, D, Lysøe, E, Gotfredsen, CH, Giese, H, and Larsen, TO. "Black perithecial pigmentation in *Fusarium* species is due to the accumulation of 5-deoxybostrycoidin-based melanin". In: *Sci Rep* 6.26206 (2016), pp. 1–13.

[10] Gatenbeck, S and Bentley, R. "Naphthaquinone biosynthesis in moulds: the mechanism for formation of javanicin". In: *Biochem J* 94.2 (1965), pp. 478–481.

[11] Arsenault, GP. "Fungal metabolites — III: Quinones from *Fusarium solani* D2 purple and structure of (+)-solaniol". In: *Tetrahedron* 24.13 (1968), pp. 4745–4749.

[12] Kurobane, I, Vining, LC, McInnes, AG, and Gerber, NN. "Metabolites of Fusarium solani related to dihydrofusarubin". In: *J Antibiot* 33.11 (1980), pp. 1376–1379.

[13] Medentsev, AG and Akimenko, VK. "Naphthoquinone metabolites of the fungi". In: *Phytochem* 47.6 (1998), pp. 935–959.

[14] Kurobane, I, Zaita, N, and Fukuda, A. "New metabolites of Fusarium martii related to dihydrofusarubin". In: *J Antibiot* 39.2 (1986), pp. 205–214.

[15] Blin, K, Wolf, T, Chevrette, MG, Lu, X, Schwalen, CJ, Kautsar, SA, Suarez Duran, HG, De Los Santos, ELC, Kim, HU, Nave, M, Dickschat, JS, Mitchell, DA, Shelest, E, Breitling, R, Takano, E, Lee, SY, Weber, T, and Medema, MH. "antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification". In: *Nucleic Acids Res* 45 (2017), pp. 36–41.

[16] Götz, S, García-Gómez, JM, Terol, J, Williams, TD, Nagaraj, SH, Nueda, MJ, Robles, M, Talón, M, Dopazo, J, and Conesa, A. "High-throughput functional annotation and data mining with the Blast2GO suite". In: *Nucleic Acids Res* 36.10 (2008), pp. 3420–3435.

[17] Sullivan, MJ, Petty, NK, and Beatson, SA. "Easyfig: a genome comparison visualizer". In: *Bioinf* 27.7 (2011), pp. 1009–1010.

[18] El-Gebali, S, Mistry, J, Bateman, A, Eddy, SR, Luciani, A, Potter, SC, Qureshi, M, Richardson, LJ, Salazar, GA, Smart, A, et al. "The Pfam protein families database in 2019". In: *Nucleic Acids Res* 47.D1 (2018), pp. D427–D432.

[19] Awakawa, T, Kaji, T, Wakimoto, T, and Abe, I. "A heptaketide naphthaldehyde produced by a polyketide synthase from *Nectria haematococca*". In: *Bioorg Med Chem Lett* 22.13 (2012), pp. 4338–4340.

[20] Graziani, S, Vasnier, C, and Daboussi, MJ. "Novel polyketide synthase from *Nectria haematococca*". In: *Appl Environ Microbiol* 70.5 (2004), pp. 2984–2988.

[21] O'Donnell, K, Rooney, AP, Proctor, RH, Brown, DW, McCormick, SP, Ward, TJ, Frandsen, RJN, Lysøe, E, Rehner, SA, Aoki, T, Robert, VA, Crous, PW, Groenewald, JZ, Kang, S, and Geiser, DM. "Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria". In: *Fungal Genet Biol* 52 (2013), pp. 20–31.

[22] Wiemann, P, Sieber, CM, Von Bargen, KW, Studt, L, Niehaus, EM, Espino, JJ, Huß, K, Michielse, CB, Albermann, S, Wagner, D, et al. "Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites". In: *PLoS Pathogens* 9.6 (2013), pp. 1–35.

[23] Parisot, D, Devys, M, and Barbier, M. "6-O-demethyl-5-deoxyfusarubin and its anhydro derivative produced by a mutant of the fungus *Nectria Haematococca* blocked in fusarubin biosynthesis". In: *J Antibiot* 44.1 (1991), pp. 103–107.

[24] Medentsev, AG, Arinbasarova, AY, and Akimenko, VK. "Biosynthesis of naphthoquinone pigments by fungi of the genus *Fusarium*". In: *Appl Biochem Microbiol* 41.5 (2005), pp. 503–507.

[25] Heiser, I, Osswald, W, Baker, R, Nemec, S, and Elstner, EF. "Activation of *Fusarium* naphthazarin toxins and other p-quinones by reduced thioctic acid". In: *J Plant Physiol* 153.3-4 (1998), pp. 276–280.

[26] Nemec, S, Phelps, D, and Baker, R. "Effects of dihydrofusarubin and isomarticin from *Fusarium solani* on carbohydrate status and metabolism of rough lemon seedlings". In: *Phytopathol* 79.6 (1989), pp. 700–705.

[27] Parisot, D, Maugin, M, and Gerlinger, C. "Genetic and epigenetic factors involved in the excretion of naphthoquinone pigments into the culture medium by *Nectria haematococca*". In: *Microbiol* 126.2 (1981), pp. 443–457.

[28] Tatum, JH and Baker, RA. "Naphthoquinones produced by Fusarium solani isolated from citrus". In: *Phytochem* 22.2 (1983), pp. 543–547.

[29] Marcinkowska, J, Kraft, JM, and Marquis, LY. "Phytotoxic effects of cell-free cultural filtrates of *Fusarium solani* isolates on virulence, host specificity and resistance". In: *Can J Plant Sci* 62.4 (1982), pp. 1027–1035.

[30] Parisot, D, Devys, M, and Barbier, M. "Conversion of anhydro-fusarubin lactol into the antibiotic bostrycoidin". In: *J Antibiot* 42.7 (1989), pp. 1189–1190.

[31] Chen, W, Chen, R, Liu, Q, He, Y, He, K, Ding, X, Kang, L, Guo, X, Xie, N, Zhou, Y, Lu, Y, Cox, RJ, Molnár, I, Li, M, Shao, Y, and Chen, F. "Orange, red, yellow: biosynthesis of azaphilone pigments in Monascus fungi". In: *Chem Sci* 8.7 (2017), pp. 4917–4925.

[32] van Wagoner, RM, Mantle, PG, and Wright, JLC. "Biosynthesis of scorpinone, a 2-azaanthraquinone from Amorosia littoralis, a fungus from marine sediment". In: *J Nat Prod* 71.3 (2008), pp. 426–430.

[33] Parisot, D, Devys, M, and Barbier, M. "Naphthoquinone pigments related to fusarubin from the fungus *Fusarium solani* (Mart.) Sacc." In: *Microbios* 64 (1990), pp. 31–47.

[34] Parisot, D, Devys, M, and Barbier, M. "Structure and biosynthesis of 5-deoxyfusarubin and anhydro-5-deoxyfusarubin, naphthaquinone pigments from Nectria haematococca". In: *Phytochem* 24.9 (1985), pp. 1977–1979.

[35] Tatum, JH, Baker, RA, and Berry, RE. "Metabolites of *Fusarium solani*". In: *Phytochem* 28.1 (1989), pp. 283–284.

[36] Tatum, JH, Baker, RA, and Berry, RE. "Naphthoquinones produced by *Fusarium oxysporum* isolated from citrus". In: *Phytochem* 24.3 (1985), pp. 457–459.

[37] Kern, H and Naef-Roth, S. "Zwei neue, durch Martiella-Fusarien gebildete Naphthazarin-Derivate 1". In: *J Phytopathol* 60.4 (1967), pp. 316–324.

[38] Atanasova-Penichon, V, Legoahec, L, Bernillon, S, Deborde, C, Maucourt, M, Verdal-Bonnin, MN, Pinson-Gadais, L, Ponts, N, Moing, A, and Richard-Forget, F. "Mycotoxin biosynthesis and central metabolism are two interlinked pathways in *Fusarium graminearum*, as demonstrated by the extensive metabolic changes induced by caffeic acid exposure". In: *Appl Environ Microbiol* 84.8 (2018), pp. 1–20.

[39] Holenstein, JE, Kern, H, Stoessl, A, and Stothers, JB. "The marticins: evidence for a mixed origin from the polyketide and tricarboxylic acid pathways by [2-$^{13}C_1$] and [1, 2-$^{13}C_2$]-acetate incorporation experiments". In: *Tetrahedron Lett* 24.38 (1983), pp. 4059–4062.

[40] Kimura, Y, Hamasaki, T, and Nakajima, H. "Isolation, identification and biological activities of 8-O-methyl-javanicin produced by *Fusarium solani*". In: *Agric Biol Chem* 45.11 (1981), pp. 2653–2654.

[41] Tawfik, OK and S, D. "Enzyme promiscuity: a mechanistic and evolutionary perspective". In: *Annu Rev Biochem* 79 (2010), pp. 471–505.

[42] Dewick, PM. *Medicinal Natural Products - A Biosynthetic Approach, Third Edition.* 3rd. John Wiley & Sons Ltd, 2009.

6 Overall discussion, conclusion and perspectives

The overall goal for this PhD project was to develop a platform for design and production of cyclic and aromatic polyketides. This goal was addressed in chapter 3 and 4 with two manuscripts describing a proof-of-concept for the general platform for designing polyketides with a given chain length and folding pattern, and the application of the platform to produce a compound of interest, respectively. Finally, chapter 5 gave an example of how nature generates a large chemical diversity starting from one common polyketide scaffold.

Chapter 3 - Manuscript I found that it was possible to produce an array of polyketides (C_6 to C_{16}) produced by heterologous expression of type III PKSs in *Saccharomyces cerevisiae*. Furthermore, we showed that control of the folding of the polyketide backbone was possible by co-expression of type II PKS cyclases, which directed either the C7-C12 or C9-C14 folding patterns. The results presented in the manuscript offers a proof-of-concept for the platform and lays the foundation for further work to develop the potential of the platform. As described in chapter 2 the longest chain produced by a type III PKS was by the engineered AaOKS, which with two point mutations in the internal catalytic chamber allowed the OKS to produce a dodecaketide (C_{24}) chain [1]. The formation of longer chains from type III PKSs may be possible to achieve through further protein engineering of the OKS or other type III PKSs, as well as investigation of natural sources for new type III PKSs. If the platform was able to create longer polyketide chains than the current C_{16} length, the product range would enter into some very interesting compounds of pharmaceutical importance. e.g. tetracyclines [2].

Another strong improvement of the platform would be to investigate other cyclases. The current C7-C12 cyclase gra-orf4 was not very efficient and much of the produced polyketide ended up as the shunt products SEK4b and dSEK4b. This issue could be alleviated by testing other C7-C12 promoting cyclases. An initial test would be ZhuI, which was implemented in the flavokermesic acid producing strain in manuscript II, as ZhuI improved the C7-C12 percentage from ∼30% to ∼70% compared to the ∼48% to ∼54% improvement by gra-orf4 in manuscript

I. This highlights that the performance of cyclase enzymes varies substantially. Therefore, the next step to develop the platform would be to test other cyclases promoting the C7-C12 or C9-C14 folding pattern, and benchmark these against the already tested cyclases. With the first ring folding being well established, further ring formations could also be cyclase catalyzed by implementation of more cyclases, as exemplified in manuscript II. The number of possible folding patterns rise substantially when directing more ring formations and the number of possible compounds would likewise increase. Finally, it would be important to test substrate specificity and limitation of the individual cyclases as different cyclases may not be able to catalyze folding of the same polyketide chain length.

When comparing the performance of OKS expressed alone in manuscript I and manuscript II, the level of produced metabolites and the percentages of the different folding patterns differ, with manuscript I and II having a C7-C12% of ∼48% and ∼35% respectively. Furthermore, standalone expression of OKS in manuscript II produced flavokermesic acid, while this was not found in detectable amounts in manuscript I. To offer an explanation for this it is important to take the difference in expression strategies of the two studies into consideration. In manuscript I the OKS was expressed on a 2μ high copy plasmid, which ensured a high level of expression, while expression in manuscript II was by integration into the genome of *S. cerevisiae*, which has lower expression levels [3]. Furthermore, the cultivation medium in manuscript I was the minimal medium synthetic complete, while the cultivation medium in manuscript II was the complex medium YPD, which changes the metabolic state of yeast [4]. The discrepancies in OKS performance can therefore be attributed to different levels of OKS enzyme being formed and the metabolic state of the yeast cells in the different media. The 2μ plasmid is a good tool for discovery as it offers high gene expression while it is not applicable for an industrial cell factory. For cell factory design integration into the genome is more applicable and other methods such as integration of multiple copies of the heterologous gene or genes can be used for increasing gene expression [5].

Chapter 4 - Manuscript II applied the principle of the platform to produce a polyketide product of interest, flavokermesic acid which is a precursor to the industrial pigment carminic acid. The artificial biosynthetic pathway for flavokermesic acid production had previously been expressed in *Nicotiana benthamiana* [6] and *Aspergillus nidulans* [7]. Furthermore, the three involved enzymes were fused in all possible configurations in an attempt to improve production of the pathway.

However, it was found that the highest production of flavokermesic acid was from the strain expressing the three enzymes as non-fused proteins yielding 52 mg flavokermesic acid per liter of culture. This showed that the fusion method applied here was not appropriate for the given pathway. This finding was attributed to the fact that the biosynthetic enzymes are likely homodimers. Alternative design of linkers with them being longer and/or less flexible to create distance between the biosynthetic enzymes could be tested [8, 9], while An alternative method to bring biosynthetic enzymes in close vicinity could be enzyme scaffolding [10, 11]. The utilization of scaffolds and peptide ligands instead of translational enzyme fusion to bring biosynthetic pathway enzymes in close vicinity could alleviate the issues with enzymes being unable to form dimers and allow dimers to form. This method of scaffolding could also give a better understanding of the correct ratio between enzymes needed for optimal production of a desired product [10].

Another factor to optimize for increased flavokermesic acid production would be expression of biosynthetic enzymes in a yeast strain that has been optimized for production of polyketides. Many steps can be taken towards improving a *S. cerevisiae* strain for production of polyketides. This could be increasing the cytosolic pool of malonyl-CoA to improve substrate availability for polyketide production [12, 13, 14], considering compartmentalisation [15, 16] or amplification of the pathway [5], and fine tuning of enzymatic levels to maximize output of the biosynthetic pathway [3, 17, 18].

Two more biosynthetic steps are needed to convert flavokermesic acid to carminic acid. One being the specific oxidation and another being glycosylation. The latter of the two reactions has been found to be catalyzed by UGT2 from *Dactylopius coccus*, while a P450 monooxygenase able to catalyze the oxidation is currently unknown. Kermesic acid and carminic acid both having the needed oxidation was detected in *A. nidulans* when OKS, ZhuI, ZhuJ and UGT2 were co-expressed, indicating that the oxidation was carried out by an *A. nidulans* endogenous monooxyganse [7]. If a suitable monooxygenase was found and co-expressed with the biosynthetic enzymes it may be possible to produce carminic acid in *S. cerevisiae*.

Chapter 5 - Manuscript III investigated the biosynthetic gene cluster *pgl1*, which is widely distributed in the fungal genus *Fusarium*. Furthermore, previously isolated and described compounds assumed to originate from the common *pgl1* biosynthetic gene cluster, were put into one biosynthetic context named a metabiosynthetic pathway. The biosynthetic gene cluster was present in all but

four species complexes of the *Fusarium* genus, which is in contrast to previous studies, which found it in all investigated species [19, 20]. The species containing the cluster all had the previously identified genes present with high homology between the genes while the topology of the cluster was less conserved across the genus.

To fully understand the biosynthesis of *pgl1* biosynthetic gene cluster derived metabolites further studies are needed, as some of the proposed biosynthetic steps could not be linked to the current *pgl1* biosynthetic gene cluster. The next steps to gain insight on this would be to investigate species from the *solani* species complex as these have been a rich source for isolation of the *pgl1* biosynthetic cluster derived compounds [21].

If the chemical diversity from the *pgl1* biosynthetic gene cluster is produced mostly from the current recognized genes, the *pgl1* metabiosynthetic pathway serves as an example of how few enzymes can bring about very high levels of chemodiversity. This is hypothesized to occur because the enzymes are promiscuous in their substrate and reaction specificity. The concept of having one pathway producing many different compounds can be an advantage over producing one or few compounds, as chemodiversity also lead to diversity in bioactivity, and thus likely a competitive advantage for the producing organism [22].

The chemodiversity originating from few biosynthetic enzymes as described in manuscript III can also be used in the context of the platform in manuscript I. The chemodiversity from relatively few enzymes can be utilized to create a broad array of compounds. If the platform was to be used for drug development it would be desirable to be able to produce a wide array of compounds in the sense of development of novel bioactivities. Here the promiscuity of the enzymes would be a positive trait [23]. On the other hand, this would be undesirable if the platform was to be utilized to create cell factories. For creation of a cell factory to produce one or very few products the involved enzymes have to be very specific and efficient [24]. The specificity of enzymes can be increased by enzyme engineering or directed evolution strategies [25]. With that being said it may be more giving to search for enzymes with higher specificity than utilizing the genes found in the *pgl1* biosynthetic gene cluster for creation of a cell factory.

Overall, the results presented in this thesis presents an approach to produce polyketides in a programmable fashion. The potential of the platform needs further work to be fully developed with longer polyketide chains being produced and

further direction of the folding of the polyketide backbone. Furthermore, the incorporation of tailoring enzymes to carry out specific reductions, oxidations, methylations, glycosylations *et al.* would vastly expand the chemodiversity, which the platform would be able to produce.

References

[1] Wanibuchi, K, Morita, H, Noguchi, H, and Abe, I. "Enzymatic formation of an aromatic dodecaketide by engineered plant polyketide synthase". In: *Bioorg Med Chem Lett* 21.7 (2011), pp. 2083–2086.

[2] Gomes, ES, Schuch, V, and Lemos, EGdM. "Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics". In: *Braz J Microbiol* 44.4 (2013), pp. 1007–1034.

[3] Lee, ME, DeLoache, WC, Cervantes, B, and Dueber, JE. "A highly characterized yeast toolkit for modular, multipart assembly". In: *ACS Synth Biol* 4.9 (2015), pp. 975–986.

[4] Jeffries, TW and Jin, YS. "Metabolic engineering for improved fermentation of pentoses by yeasts". In: *Appl Microbiol Biotechnol* 63.5 (2004), pp. 495–509.

[5] Strucko, T, Buron, LD, Jarczynska, ZD, Nødvig, CS, Mølgaard, L, Halkier, BA, and Mortensen, UH. "CASCADE, a platform for controlled gene amplification for high, tunable and selection-free gene expression in yeast". In: *Sci Rep* 7.41431 (2017), pp. 1–12.

[6] Andersen-Ranberg, J, Kongstad, KT, Nafisi, M, Staerk, D, Okkels, FT, Mortensen, UH, Lindberg Møller, B, Frandsen, RJN, and Kannangara, R. "Synthesis of C-Glucosylated Octaketide Anthraquinones in *Nicotiana benthamiana* by Using a Multispecies-Based Biosynthetic Pathway". In: *ChemBioChem* 18.19 (2017), pp. 1893–1897.

[7] Frandsen, RJ, Khorsand-Jamal, P, Kongstad, KT, Nafisi, M, Kannangara, RM, Staerk, D, Okkels, FT, Binderup, K, Madsen, B, Møller, BL, Thrane, U, and Mortensen, UH. "Heterologous production of the widely used natural food colorant carminic acid in *Aspergillus nidulans*". In: *Sci Rep* 8.12853 (2018), pp. 1–10.

[8] Albertsen, L, Chen, Y, Bach, LS, Rattleff, S, Maury, J, Brix, S, Nielsen, J, and Mortensen, UH. "Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes". In: *Appl Environ Microbiol* 77.3 (2011), pp. 1033–1040.

[9] Li, G, Huang, Z, Zhang, C, Dong, BJ, Guo, RH, Yue, HW, Yan, LT, and Xing, XH. "Construction of a linker library with widely controllable flexibility for fusion protein design". In: *Applied Microbiol Biotechnol* 100.1 (2016), pp. 215–225.

[10] Dueber, JE, Wu, GC, Malmirchegini, GR, Moon, TS, Petzold, CJ, Ullal, AV, Prather, KLJ, and Keasling, JD. "Synthetic protein scaffolds provide modular control over metabolic flux". In: *Nat Biotechnol* 27.8 (2009), pp. 753–759.

[11] Jepsen, MDE, Sparvath, SM, Nielsen, TB, Langvad, AH, Grossi, G, Gothelf, KV, and Andersen, ES. "Development of a genetically encodable FRET system using fluorescent RNA aptamers". In: *Nat commun* 9.18 (2018), pp. 1–10.

[12] Wattanachaisaereekul, S, Lantz, AE, Nielsen, ML, and Nielsen, J. "Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply". In: *Met Eng* 10.5 (2008), pp. 246–254.

[13] van Rossum, HM, Kozak, BU, Pronk, JT, and van Maris, AJ. "Engineering cytosolic acetyl-coenzyme A supply in *Saccharomyces cerevisiae*: pathway stoichiometry, free-energy conservation and redox-cofactor balancing". In: *Met Eng* 36 (2016), pp. 99–115.

[14] Chen, Y, Daviet, L, Schalk, M, Siewers, V, and Nielsen, J. "Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism". In: *Met Eng* 15 (2013), pp. 48–54.

[15] Liu, L, Redden, H, and Alper, HS. "Frontiers of yeast metabolic engineering: diversifying beyond ethanol and *Saccharomyces*". In: *Curr Opin Biotechnol* 24 (2013), pp. 1023–1030.

[16] Chen, AH and Silver, PA. "Designing biological compartmentalization". In: *Trends Cell Biol* 22.12 (2012), pp. 662–670.

[17] Sun, J, Shao, Z, Zhao, H, Nair, N, Wen, F, Xu, JH, and Zhao, H. "Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*". In: *Biotechnol Bioeng* 109.8 (2012), pp. 2082–2092.

[18] Harvey, CJB, Tang, M, Schlecht, U, Horecka, J, Fischer, CR, Lin, HC, Li, J, Naughton, B, Cherry, J, Miranda, M, Li, YF, Chu, AM, Hennessy, JR, Vandova, GA, Inglis, D, Aiyar, RS, Steinmetz, LM, Davis, RW, Medema, MH, Sattely, E, Khosla, C, St. Onge, RP, Tang, Y, and Hillenmeyer, ME. "HEx: A heterologous expression platform for the discovery of fungal natural products". In: *Sci Adv* 4.eaar5459 (2018), pp. 1–14.

[19] Brown, DW and Proctor, RH. "Insights into natural products biosynthesis from analysis of 490 polyketide synthases from Fusarium". In: *Fungal Genet Biol* 89 (2016), pp. 37–51.

[20] Frandsen, RJN, Rasmussen, SA, Knudsen, PB, Uhlig, S, Petersen, D, Lysøe, E, Gotfredsen, CH, Giese, H, and Larsen, TO. "Black perithecial pigmentation in *Fusarium* species is due to the accumulation of 5-deoxybostrycoidin-based melanin". In: *Sci Rep* 6.26206 (2016), pp. 1–13.

[21] Medentsev, AG and Akimenko, VK. "Naphthoquinone metabolites of the fungi". In: *Phytochem* 47.6 (1998), pp. 935–959.

[22] Fischbach, MA and Clardy, J. "One pathway, many products". In: *Nat Chem Biol* 3.7 (2007), pp. 353–355.

[23] Dandapani, S, Germain, AR, Jewett, I, le Quement, S, Marie, JC, Muncipinto, G, Duvall, JR, Carmody, LC, Perez, JR, Engel, JC, et al. "Diversity-oriented synthesis yields a new drug lead for treatment of Chagas disease". In: *ACS Med Chem Lett* 5.2 (2014), pp. 149–153.

[24] Yang, L, Lübeck, M, and Lübeck, PS. "*Aspergillus* as a versatile cell factory for organic acid production". In: *Fungal Biol Rev* 31.1 (2017), pp. 33–49.

[25] Behrendorff, JBYH, Huang, W, and Gillam, EMJ. "Directed evolution of cytochrome P450 enzymes for biocatalysis: exploiting the catalytic versatility of enzymes with relaxed substrate specificity". In: *Biochem J* 467 (2015), pp. 1–15.

7 Appendices

7.1 Appendix – Manuscript I

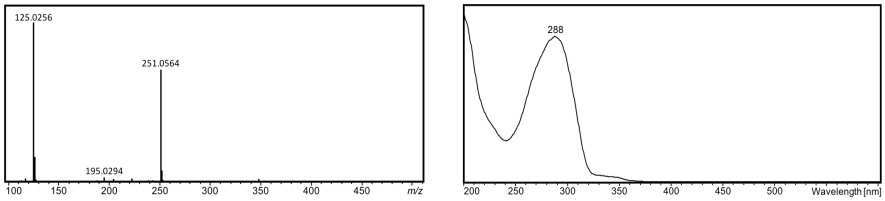
Table S3.1 List of Primers used in this study.

Name	Sequence	Purpose
Sc_Gh_2-PS-F	ATCAACGGGUAAAAATGGG TTCCTACTCTTCTGATGAT GTTG	Amplification of Gh2PS for USER cloning
Sc_Gh_2-PS-R	CGTGCGAUTTAGTTACCAT TAGCAACAGCAGCAGTAAC TC	
Sc_Aa_PCS-F	ATCAACGGGUAAAAATGTC CTCCTTGTCTAATTCCTTG C	Amplification of AaPCS for USER cloning
Sc_Aa_PCS-R	CGTGCGAUTTACATCAAAG GCAAAGAATGCA	
Sc_DluHKS-F	ATCAACGGGUAAAAATGGC TTTCGTTGAAGGTATGGGT	Amplification of DluHKS for USER cloning
Sc_DluHKS-R	CGTGCGAUTTAGTTGTTGA TTGGGAAGGATCTCAAGA	
Sc_AaPKS3-F	ATCAACGGGUAAAAATGGG TTCCTTGTCTGATTCTACT CCA	Amplification of AaPKS3 for USER cloning
Sc_AaPKS3-R	CGTGCGAUTTAGACTGGTG GCAAAGAATGCAACA	
Sc_ZhuI-F	AGCGATACGUAAAAATGAG ACACGTTGAACACACAGTT ACCG	Amplification of ZhuI for USER cloning
Sc_ZhuI-R	CACGCGAUTTATTATGCAG TTACGGTACCAACACCAC	
Sc_pdmD-F	AGCGATACGUAAAAATGAC TCAATGGAGAACCGATTCC	Amplification of pdmD for USER cloning
Sc_pdmD-R	CACGCGAUTTATCTAGCAC CTCTAGCAGCTCTTTCAA	
Sc_gra-orf4-F	AGCGATACGUAAAAATGGC TAGAACTGCTGCTTTGC	Amplification of gra-orf4 for USER cloning
Sc_gra-orf4-R	CACGCGAUTTAACCTGCTT CAGCAGCTTCAGC	
L1-pBOSAL1 FW	AGACTTCGUCCTCGTGATA CGCCTATTTTTATAG	Amplification of pBOSAL1 vector for USER cloning
L2-pBOSAL1 RV	AGTAGACTUCCCGGGAATT GCCATGAAG	
L2-pKl.LEU2 FW	AAGTCTACUAGCTCGCTGT GAAGATCC	Amplification of KlLEU2 for USER cloning
L1-tKl.LEU2 RV	ACGAAGTCUGATCCGCAGG CTAACCGG	
pBOSAL1-LEU2 SQ FW	GCTTGAGAAGGTTTTGGGA C	Sequencing of pBOSAL1-LEU2 vector
pBOSAL1-LEU2 SQ RV	TTGGATTAGTCTCATCCTT CAATG	
AaOKS FW	ATCAACGGGUAAAAATGAG TAGTTTATCAAATGCCAGT CAC	Amplification of AaOKS for USER cloning
AaOKS RV	CGTGCGAUTTACATCAATG GCAAGGAATGCAATAAG	

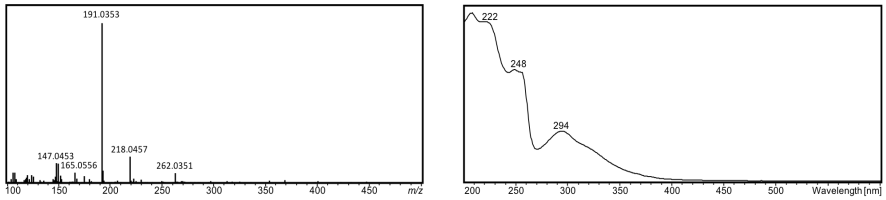
pTEF1 F	CACGCGAUCCACACACCAT AGCTTCAAAATG	Amplification of PTEF1 for USER cloning
pTEF1 R	ACCCGTTGAUTTGTAATTA AAACTTAGATTAGATTGCT ATGC	
pPGK1 F	CACGCGAUAAGTACCTTCA AAGAATGGG	Amplification of PPGK1 for USER cloning
pPGK1 R	ACCCGTTGAUTGTTTTATA TTTGTTGTAAAAAGTAGAT AATTAC	
PGK1-R	ATCATCAAGGAAGTAATTA TCT	Colony PCR / sequencing
ADH1-F	CATAAATCATAAGAAATTC GCT	
TEF1_F	ACTTTTTTTACTTCTTGCT CAT	
CYC1-R2	GGACCTAGACTTCAGGTT	
XI-2-up-out-sq	CAATATCAGTGTGGT GAAC	Colony PCR
C1_ADH1_F	CTTGAGTAACTCTTTCCTG TAGGTC	

Figure S3.1 MS and UV spectra of tentatively identified compounds from heterologous expression of type III PKSs.

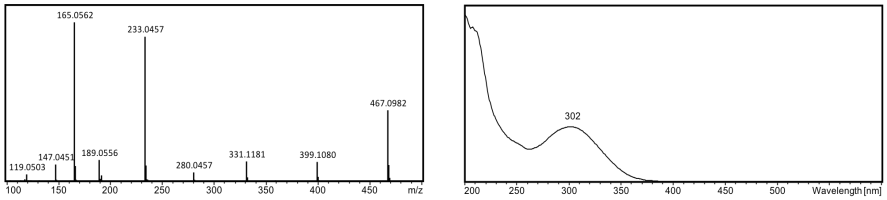
MS (ESI$^-$) and UV spectra of triacetic acid lactone (**1**).



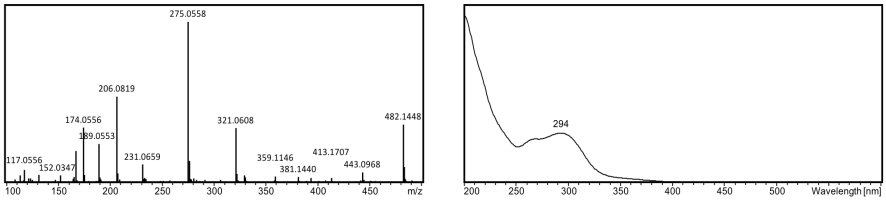
MS (ESI$^-$) and UV spectra of 5,7-dihydroxy-2-methylchromone (**3**).



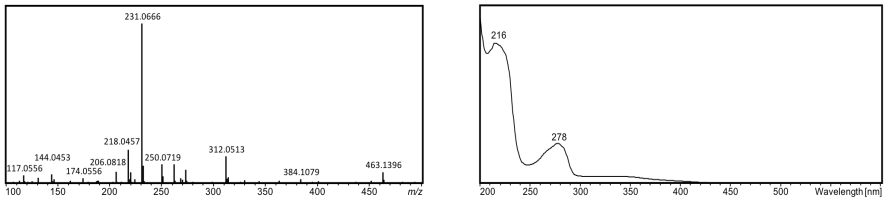
MS (ESI$^-$) and UV spectra of 6-(2',4'-dihydroxy-6'-methylphenyl)-4-hydroxy-2-pyrone (**2**).



MS (ESI$^-$) and UV spectrum of TW93a (**4**).



MS (ESI$^-$) and UV spectra of aloesone (**5**).



MS (ESI$^-$) and UV spectra of UWM1 (**11**).

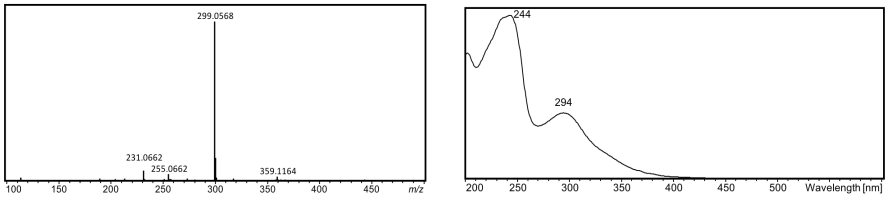
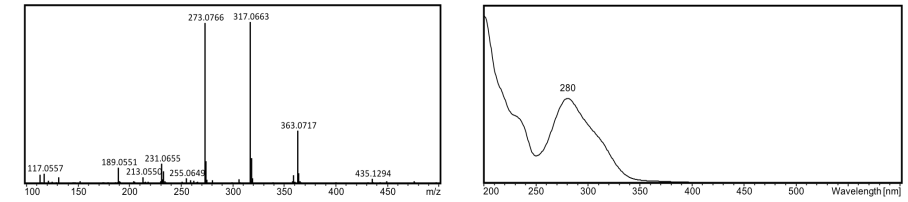
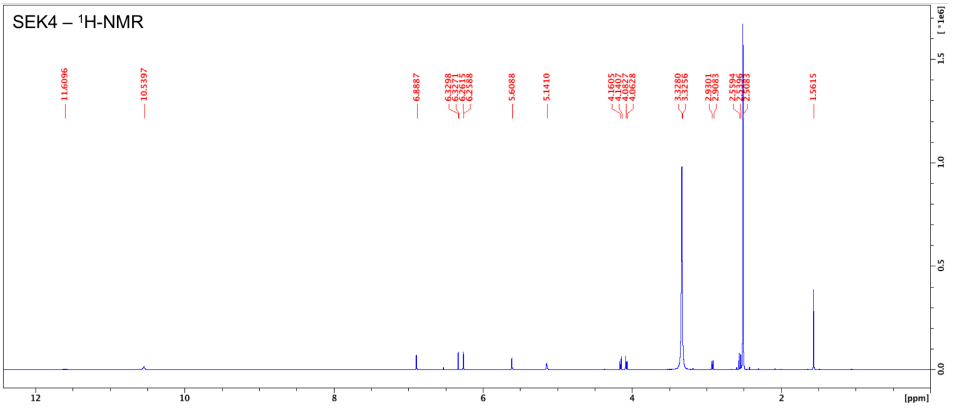
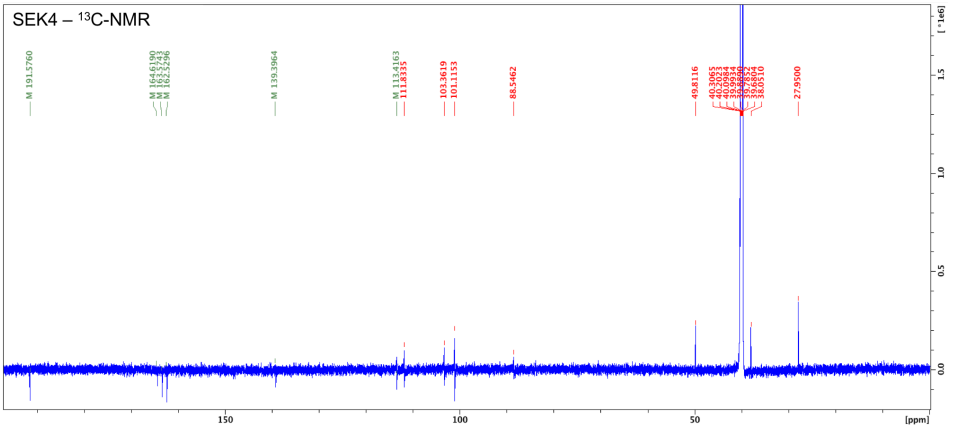
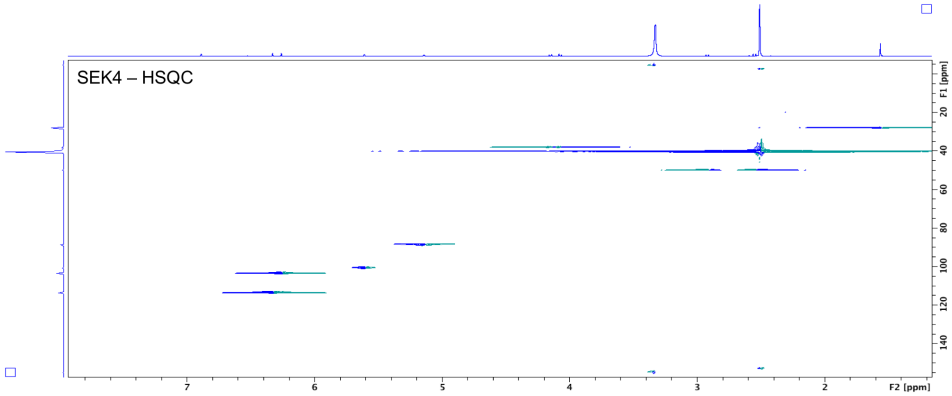


Figure S3.2 UV and MS spectra of SEK4 (**6**).








SEK4 – HSQC
F1 [ppm]
20
40
60
80
100
120
140
7
6
5
4
3
2
F2 [ppm]

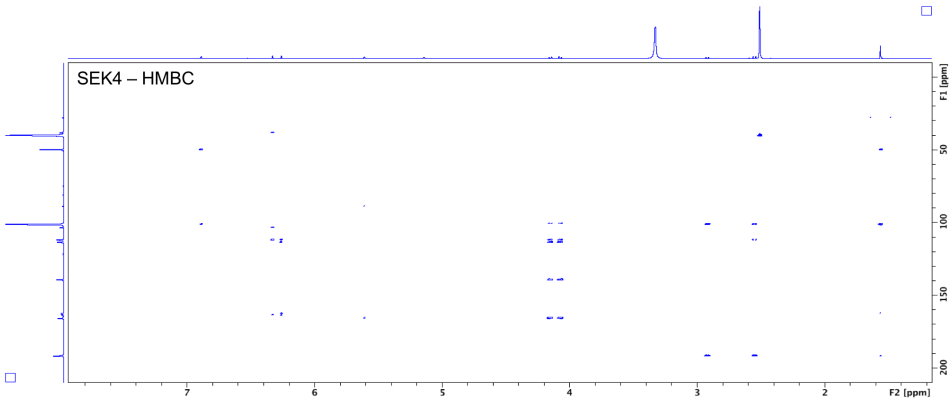
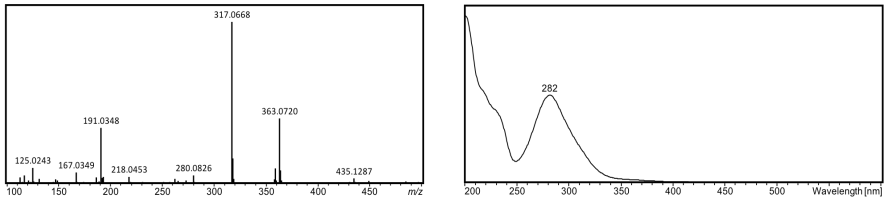
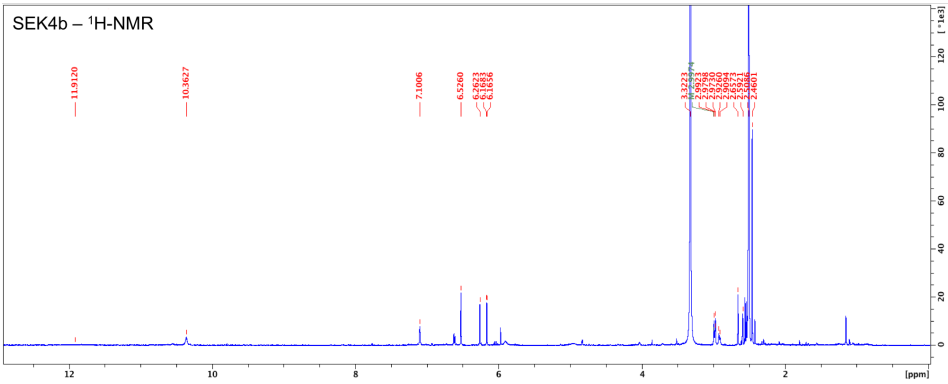
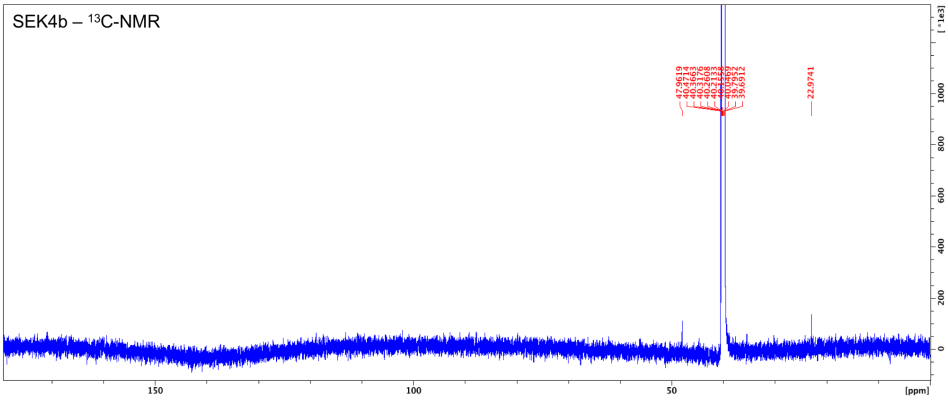
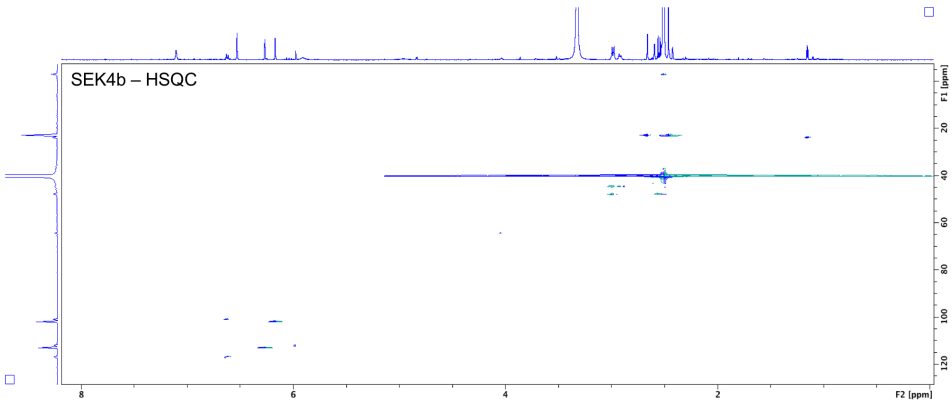

SEK4 – HMBC
F1 [ppm]
50
100
150
200
7
6
5
4
3
2
F2 [ppm]

Figure S3.3 UV and MS spectra of SEK4b (**7**).








SEK4b – HSQC
F1 [ppm]
20
40
60
80
100
120
8
6
4
2
F2 [ppm]

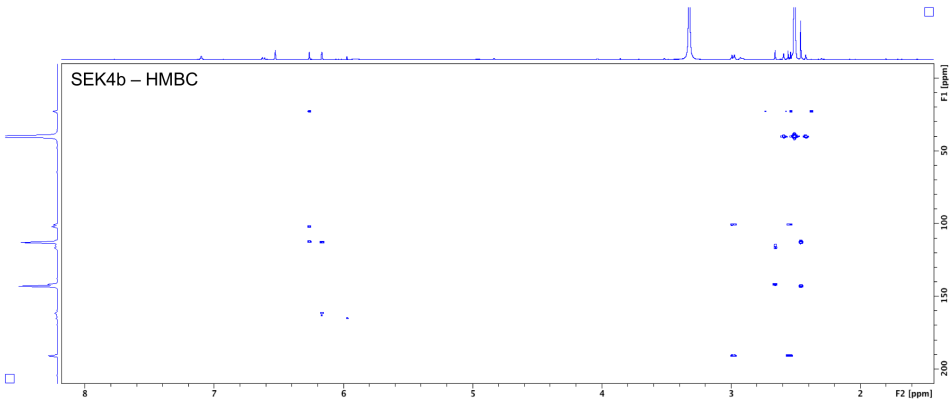
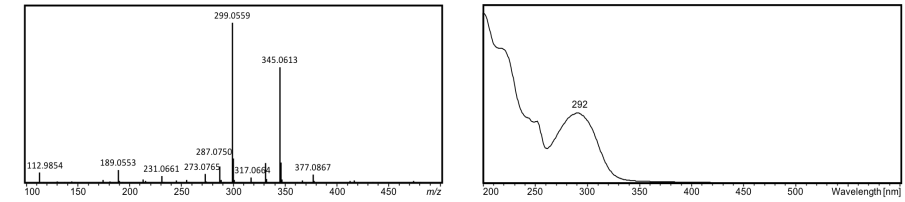
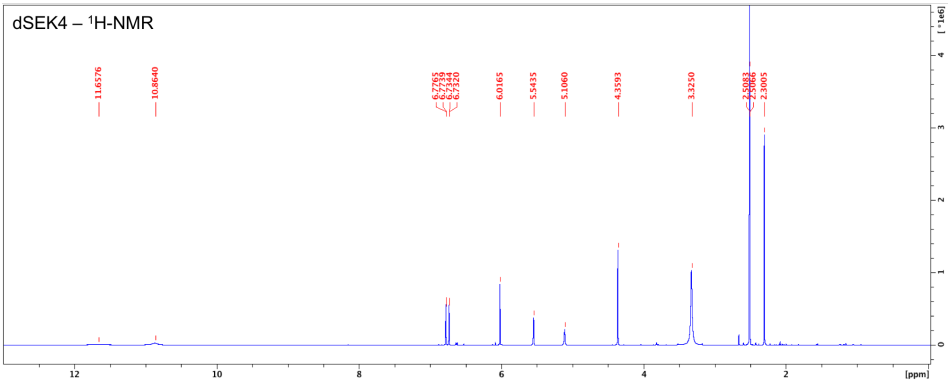
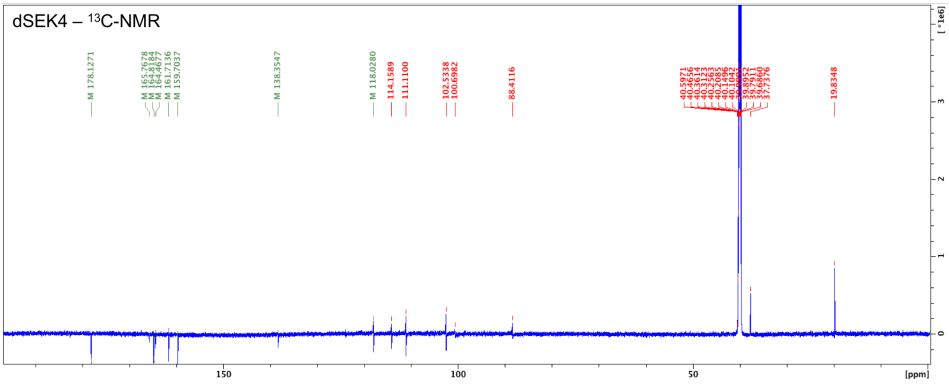
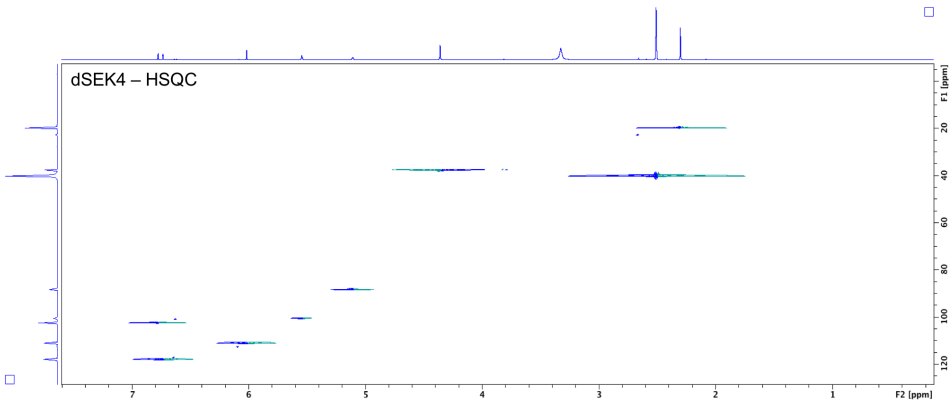

SEK4b – HMBC
F1 [ppm]
50
100
150
200
8
7
6
5
4
3
2
F2 [ppm]

Figure S3.4 UV and MS spectra of dSEK4b (**8**).








dSEK4 – HSQC
F1 [ppm]
F2 [ppm]

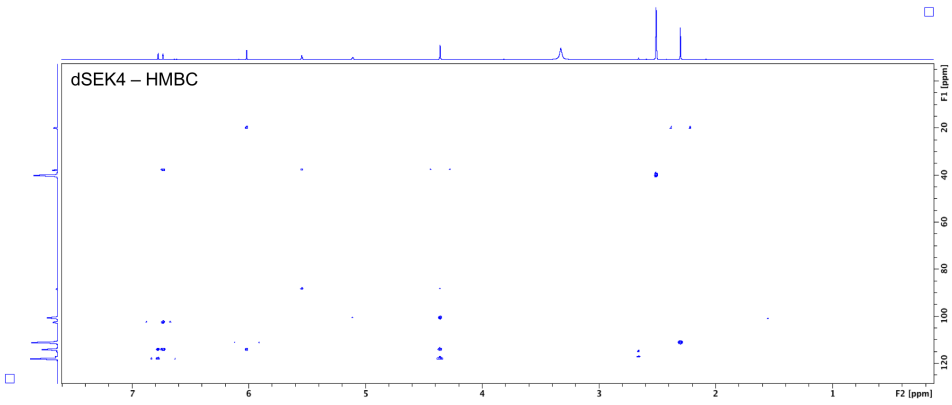
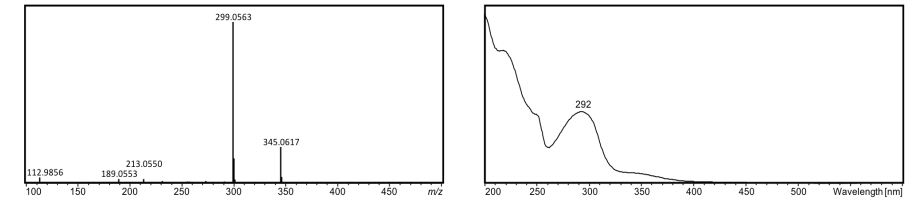
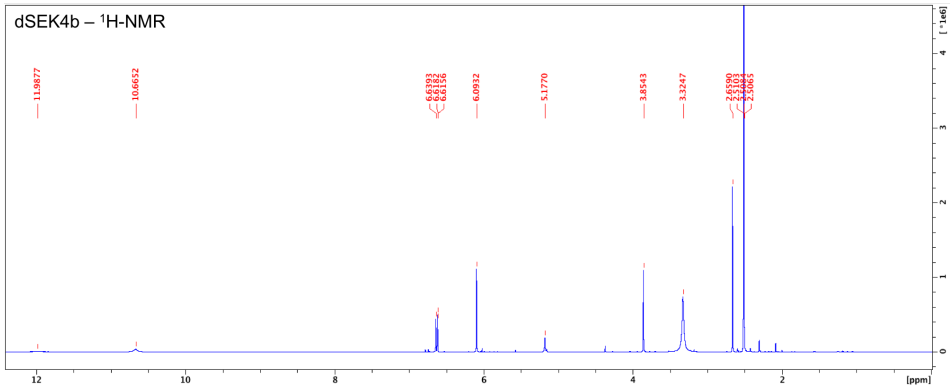
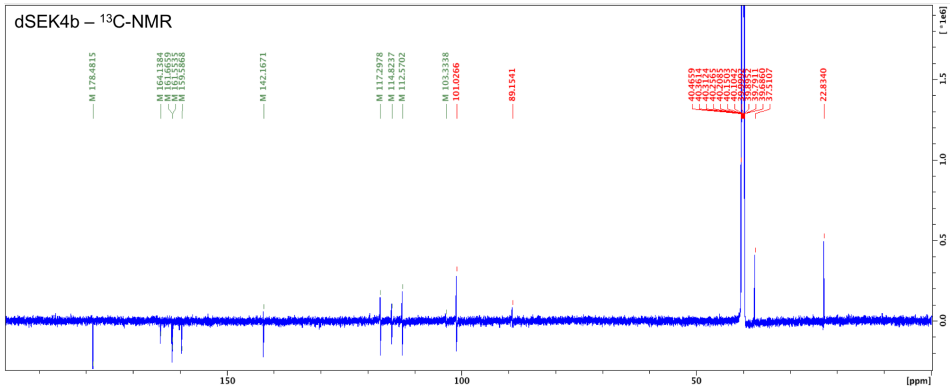
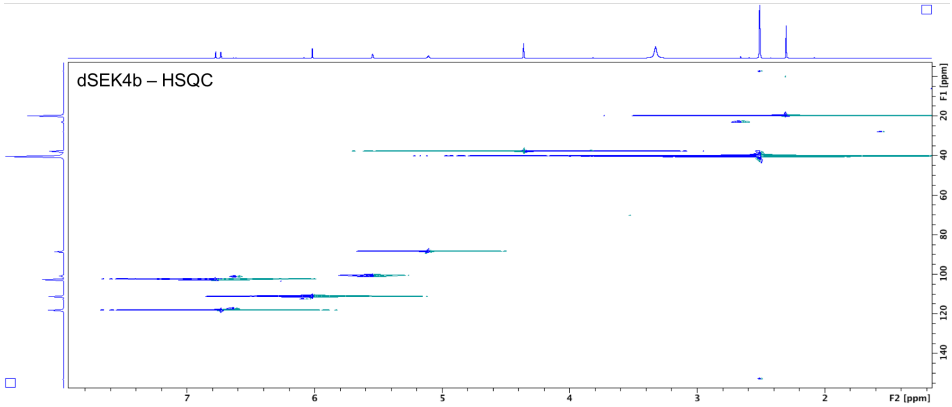

dSEK4 – HMBC
F1 [ppm]
F2 [ppm]

Figure S3.5 UV and MS spectra of dSEK4 (**9**).








dSEK4b – HSQC
F1 [ppm]
20
40
60
80
100
120
140
7
6
5
4
3
2
F2 [ppm]

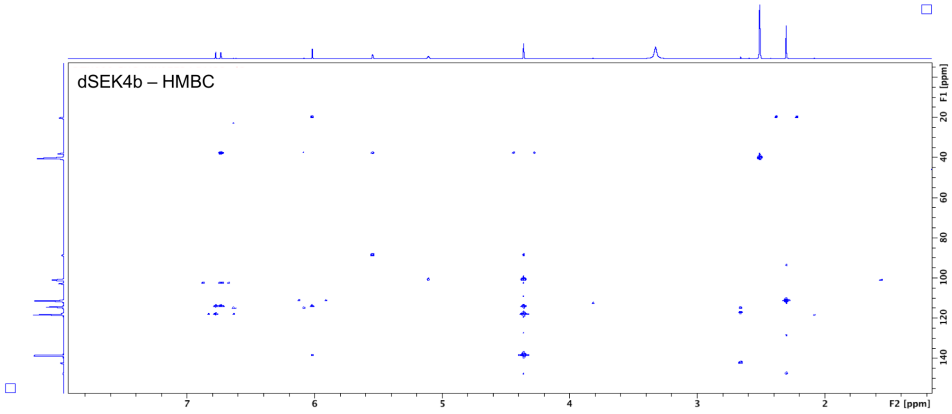

dSEK4b – HMBC
F1 [ppm]
20
40
60
80
100
120
140
7
6
5
4
3
2
F2 [ppm]

Table S3.2 The measured final OD $_{600}$ and area under peak for each of the cultivations of diploid strains.

Enzymes	OD	SEK4	SEK4b	dSEK4b	dSEK4	UWM1
AaOKS 1	9,92	4,77E+06	5,13E+06	2,75E+06	2,75E+06	1,47E+05
AaOKS 2	9,8	3,72E+06	3,89E+06	1,63E+06	1,86E+06	1,08E+05
AaOKS 3	9,88	5,02E+06	5,90E+06	3,19E+06	2,66E+06	2,32E+05
AaOKS + Gra-orf4 1	8,72	2,47E+06	1,96E+06	1,18E+06	1,37E+06	1,67E+04
AaOKS + Gra-orf4 2	7,92	2,58E+06	2,18E+06	1,13E+06	1,23E+06	1,44E+04
AaOKS + Gra-orf4 3	8,84	2,95E+06	2,42E+06	1,53E+06	1,66E+06	1,04E+04
AaOKS + PdmD 1	9,56	2,59E+06	3,11E+06	1,42E+06	1,51E+06	2,97E+06
AaOKS + PdmD 2	11,44	4,18E+06	5,26E+06	2,75E+06	2,57E+06	4,08E+06
AaOKS + PdmD 3	10,2	4,15E+06	5,34E+06	2,81E+06	2,60E+06	3,82E+06

7.2 Appendix – Manuscript II

Table S4.1 List of Primers used and constructed in this study.

Name	Sequence	Purpose
TEF1-F	CGTGCGAUGCACACACCATAGCTTCAAAATGTT	Amplification of P_{TEF1} and P_{PGK1} for USER cloning
TEF1-R	ATGACAGAUGTGAGTCGTATTACGGATCCTTG	
PGK1-R	AGTGCAGGUAAAAATGTCCGGTAG	
OKS-B-F	ATCTGTCAUAAAAATGAGTAGTTTATCAAATGCC AGTC	Amplification of OKS in first position for USER cloning
OKS-B-R	ACCACCUGAACCCATCAATGGCAAGGAATGC	
OKS-M-F	AGGTGGUGGTGGTTCTATGAGTAGTTTATCAAAT GCCAGTC	Amplification of OKS in second position for USER cloning
OKS-M-R	AGAACCUGAACCACCCATCAATGGCAAGGAATGC	
OKS-E-F	AGGTTCUGCTGGTATGAGTAGTTTATCAAATGCC AGTC	Amplification of OKS in third position for USER cloning
OKS-E-R	CACGCGAUTCACATCAATGGCAAGGAATGC	
L1-OKS-F	ATCTGTCAUAAAAATGGGTTCAGGTGGTGGTGGT TCTATGAGTAGTTTATCAAATGCCAGTC	Amplification of OKS with L1 linker for USER cloning
OKS-L1-R	CACGCGAUTCAAGAACCACCACCACCTGAACCCA TCAATGGCAAGGAATGC	
L2-OKS-F	ATCTGTCAUAAAAATGGGTGGTTCAGGTTCTGCT GGTATGAGTAGTTTATCAAATGCCAGTC	Amplification of OKS with L2 linker for USER cloning
OKS-L2-R	CACGCGAUTCAACCAGCAGAACCTGAACCACCCA TCAATGGCAAGGAATGC	
ZhuI-B-F	ATCTGTCAUAAAAATGAGACACGTTGAACACACA GTTA	Amplification of ZhuI in first position for USER cloning
ZhuI-B-R	ACCACCUGAACCTGCAGTTACGGTACCAACACCA	
ZhuI-M-F	AGGTGGUGGTGGTTCTATGAGACACGTTGAACAC ACAGTTA	Amplification of ZhuI in second position for USER cloning
ZhuI-M-R	AGAACCUGAACCACCTGCAGTTACGGTACCAACA CCA	
ZhuJ-M-F	AGGTGGUGGTGGTTCTATGTCCGGTAGAAAGACC TTTTTAG	Amplification of ZhuI in third position for USER cloning
ZhuI-E-R	CACGCGAUTCATGCAGTTACGGTACCAACACCA	
ZhuJ-B-F	ATCTGTCAUAAAAATGTCCGGTAGAAAGACCTTT TTAG	Amplification of ZhuJ in first position for USER cloning
ZhuJ-B-R	ACCACCUGAACCATCTTCTTCTTCTTGTTCGAAA ACAGC	
ZhuJ-M-F	AGGTGGUGGTGGTTCTATGTCCGGTAGAAAGACC TTTTTAG	Amplification of ZhuJ in second position for USER cloning
ZhuJ-M-R	AGAACCUGAACCACCATCTTCTTCTTCTTGTTCG AAAACAGC	
ZhuJ-E-F	AGGTTCUGCTGGTATGTCCGGTAGAAAGACCTTT TTAG	Amplification of ZhuJ in third position for USER cloning
ZhuJ-E-R	CACGCGAUTCAATCTTCTTCTTCTTGTTCGAAAA CAGC	
ZhuJ-PGK-F	AGTGCAGGUAAAAATGTCCGGTAG	Amplification of ZhuJ with the PKG1 promoter
ZhuJ-PGK-R	CGTGCGAUTCATGCAGTTACGGTACCAACACCA	
Val-OKS-Fw	CTTGCGGTGTCGATATGCCT	position for USER cloning
Val-OKS-Rv	AGGCATATCGACACCGCAAG	
Val-ZhuI-Fw	TAGGTCATATGTCTGGTGAATGGAG	
Val-ZhuI-Rv	CTCCATTCACCAGACATATGACCTA	Colony PCR / sequencing
Val-ZhuJ-Fw	GTTACGGTCCTGGTGTTAGATTG	
Val-ZhuJ-Rv	CAATCTAACACCAGGACCGTAAC	
ADH1-sq-F	CATAAATCATAAGAAATTCGCT	
CYC1-sq-R	GGACCTAGACTTCAGGTT	

Table S4.2 List of vectors and *S. cerevisiae* strains used and constructed in this study. All vectors carry the Kl.URA3 selectable marker, and all strains except the starting CEN.PK 113-5B strain are uracil prototroph.

Name	Description	Source
pX-3	X-3	[1]
pXI-2	XI-2	[1]
pFusP01	X-3::P_{TEF1}-OKS	This study
pFusP02	X-3::P_{TEF1}-OKS-L1	This study
pFusP03	X-3::P_{TEF1}-L1-OKS	This study
pFusP04	X-3::P_{TEF1}-OKS-L2	This study
pFusP05	X-3::P_{TEF1}-L2-OKS	This study
pFusP06	XI-2::P_{TEF1}-ZhuI	This study
pFusP07	XI-2::P_{TEF1}-ZhuJ	This study
pFusP08	XI-2::P_{PGK1}ZhuJ;P_{TEF1}-ZhuI	This study
pFusP09	X-3::P_{TEF1}-OKS-L1-ZhuI	This study
pFusP10	X-3::P_{TEF1}-ZhuI-L1-OKS	This study
pFusP11	X-3::P_{TEF1}-OKS-L2-ZhuI	This study
pFusP12	X-3::P_{TEF1}-ZhuI-L2-OKS	This study
pFusP13	X-3::P_{TEF1}-OKS-L1-ZhuJ	This study
pFusP14	X-3::P_{TEF1}-ZhuJ-L1-OKS	This study
pFusP15	X-3::P_{TEF1}-OKS-L2-ZhuJ	This study
pFusP16	X-3::P_{TEF1}-ZhuJ-L2-OKS	This study
pFusP17	XI-2::P_{TEF1}-ZhuI-L1-ZhuJ	This study
pFusP18	XI-2::P_{TEF1}-ZhuI-L2-ZhuJ	This study
pFusP19	XI-2::P_{TEF1}-ZhuJ-L1-ZhuI	This study
pFusP20	XI-2::P_{TEF1}-ZhuJ-L2-ZhuI	This study
pFusP21	X-3::P_{TEF1}-OKS-L1-ZhuI-L2-ZhuJ	This study
pFusP22	X-3::P_{TEF1}-OKS-L1-ZhuJ-L2-ZhuI	This study
pFusP23	X-3::P_{TEF1}-ZhuI-L1-OKS-L2-ZhuJ	This study
pFusP24	X-3::P_{TEF1}-ZhuJ-L1-OKS-L2-ZhuI	This study
pFusP25	X-3::P_{TEF1}-ZhuI-L1-ZhuJ-L2-OKS	This study
pFusP26	X-3::P_{TEF1}-ZhuJ-L1-ZhuI-L2-OKS	This study
pFusP27	XI-2::P_{TEF1}-L1-ZhuI	This study
pFusP28	XI-2::P_{TEF1}-ZhuI-L1	This study
pFusP29	XI-2::P_{TEF1}-L2-ZhuI	This study
pFusP30	XI-2::P_{TEF1}-ZhuI-L2	This study
pFusP31	XI-2::P_{TEF1}-L1-ZhuI-L2	This study
pFusP32	XI-2::P_{TEF1}-L1-ZhuJ	This study
pFusP33	XI-2::P_{TEF1}-ZhuJ-L1	This study
pFusP34	XI-2::P_{TEF1}-L2-ZhuJ	This study
pFusP35	XI-2::P_{TEF1}-ZhuJ-L2	This study
pFusP36	XI-2::P_{TEF1}-L1-ZhuJ-L2	This study

Table S4.3 List of *S. cerevisiae* strains used and constructed in this study. All strains except the starting CEN.PK 113-5B strain were uracil prototroph. The Kl.URA3 marker was in X-3 if only one insertion was done or in XI-2 if two gene insertions were done. All strains are derived from the CEN.PK 113-5B.

Name	Genotype	Reference
	Strains expressing OKS only	
CEN.PK 113-5B	*MATa MAL2-8c SUC2 ura3-52*	[2]
FusP-SC01	*X-3::*P_{TEF1}*-OKS*	This study
FusP-SC02	*X-3::*P_{TEF1}*-OKS-L1*	This study
FusP-SC03	*X-3::*P_{TEF1}*-L1-OKS*	This study
FusP-SC04	*X-3::*P_{TEF1}*-OKS-L2*	This study
FusP-SC05	*X-3::*P_{TEF1}*-L2-OKS*	This study
FusP-SC06	*X-3::*P_{TEF1}*-OKS XI-2::(Empty)*	This study
	Strains expressing OKS and one cyclase	
FusP-Sc07	*X-3::OKS, XI-2::ZhuI*	This study
FusP-Sc08	*X-3::OKS, XI-2::L1-ZhuI*	This study
FusP-Sc09	*X-3::OKS, XI-2::ZhuI-L1*	This study
FusP-Sc10	*X-3::OKS, XI-2::L2-ZhuI*	This study
FusP-Sc11	*X-3::OKS, XI-2::L1-ZhuI-L2*	This study
FusP-Sc12	*X-3::OKS-L1-ZhuI*	This study
FusP-Sc13	*X-3::ZhuI-L1-OKS*	This study
FusP-Sc14	*X-3::OKS-L2-ZhuI*	This study
FusP-Sc15	*X-3::ZhuI-L2-OKS*	This study
FusP-Sc16	*X-3::OKS, XI-2::ZhuJ*	This study
FusP-Sc17	*X-3::OKS, XI-2::L1-ZhuJ*	This study
FusP-Sc18	*X-3::OKS, XI-2::ZhuJ-L1*	This study
FusP-Sc19	*X-3::OKS, XI-2::L2-ZhuJ*	This study
FusP-Sc20	*X-3::OKS, XI-2::L1-ZhuJ-L2*	This study
FusP-Sc21	*X-3::OKS-L1-ZhuJ*	This study
FusP-Sc22	*X-3::ZhuJ-L1-OKS*	This study
FusP-Sc23	*X-3::OKS-L2-ZhuJ*	This study
FusP-Sc24	*X-3::ZhuJ-L2-OKS*	This study
	Strains expressing OKS and two cyclases	
FusP-SC25	*X-3::*P_{TEF1}*-OKS XI-2::*P_{PGK1}*-ZhuJ;*P_{TEF1}*-ZhuI*	This study
FusP-SC26	*X-3::*P_{TEF1}*-OKS XI-2::*P_{TEF1}*-ZhuI-L1-ZhuJ*	This study
FusP-SC27	*X-3::*P_{TEF1}*-OKS XI-2::*P_{TEF1}*-ZhuI-L2-ZhuJ*	This study
FusP-SC28	*X-3::*P_{TEF1}*-OKS XI-2::*P_{TEF1}*-ZhuJ-L1-ZhuI*	This study
FusP-SC29	*X-3::*P_{TEF1}*-OKS XI-2::*P_{TEF1}*-ZhuJ-L2-ZhuI*	This study
FusP-SC30	*X-3::*P_{TEF1}*-OKS-L1-ZhuI XI-2::*P_{TEF1}*-ZhuJ*	This study
FusP-SC31	*X-3::*P_{TEF1}*-ZhuI-L1-OKS XI-2::*P_{TEF1}*-ZhuJ*	This study
FusP-SC32	*X-3::*P_{TEF1}*-OKS-L2-ZhuI XI-2::*P_{TEF1}*-ZhuJ*	This study
FusP-SC33	*X-3::*P_{TEF1}*-ZhuI-L2-OKS XI-2::*P_{TEF1}*-ZhuJ*	This study
FusP-SC34	*X-3::*P_{TEF1}*-OKS-L1-ZhuJ XI-2::*P_{TEF1}*-ZhuI*	This study
FusP-SC35	*X-3::*P_{TEF1}*-ZhuJ-L1-OKS XI-2::*P_{TEF1}*-ZhuI*	This study
FusP-SC36	*X-3::*P_{TEF1}*-OKS-L2-ZhuJ XI-2::*P_{TEF1}*-ZhuI*	This study
FusP-SC37	*X-3::*P_{TEF1}*-ZhuJ-L2-OKS XI-2::*P_{TEF1}*-ZhuI*	This study
FusP-SC38	*X-3::*P_{TEF1}*-OKS-L1-ZhuI-L2-ZhuJ*	This study
FusP-SC39	*X-3::*P_{TEF1}*-OKS-L1-ZhuJ-L2-ZhuI*	This study
FusP-SC40	*X-3::*P_{TEF1}*-ZhuI-L1-OKS-L2-ZhuJ*	This study
FusP-SC41	*X-3::*P_{TEF1}*-ZhuJ-L1-OKS-L2-ZhuI*	This study
FusP-SC42	*X-3::*P_{TEF1}*-ZhuI-L1-ZhuJ-L2-OKS*	This study
FusP-SC43	*X-3::*P_{TEF1}*-ZhuJ-L1-ZhuI-L2-OKS*	This study
	Strains not expressing OKS	
FusP-SC44	*XI-2::*P_{TEF1}*-ZhuI*	This study
FusP-SC45	*XI-2::*P_{TEF1}*-ZhuJ*	This study
FusP-SC46	*XI-2::*P_{PGK1}*ZhuJ-*P_{TEF1}*-ZhuI*	This study
FusP-SC47	*X-3::(Empty) XI-2::*P_{TEF1}*-ZhuI*	This study
FusP-SC48	*X-3::(Empty) XI-2::*P_{TEF1}*-ZhuJ*	This study
FusP-SC49	*X-3::(Empty) XI-2::*P_{PGK1}*ZhuJ/*P_{TEF1}*-ZhuI*	This study
FusP-SC50	*XI-2::*P_{TEF1}*-L1-ZhuI*	This study
FusP-SC51	*XI-2::*P_{TEF1}*-ZhuI-L1*	This study
FusP-SC52	*XI-2::*P_{TEF1}*-L2-ZhuI*	This study
FusP-SC53	*XI-2::*P_{TEF1}*-L1-ZhuI-L2*	This study
FusP-SC54	*XI-2::*P_{TEF1}*-L1-ZhuJ*	This study
FusP-SC55	*XI-2::*P_{TEF1}*-ZhuJ-L1*	This study
FusP-SC56	*XI-2::*P_{TEF1}*-L2-ZhuJ*	This study

FusP-SC57	*XI-2::*P_{TEF1}*-L1-ZhuJ-L2*	This study
FusP-SC58	*X-3::(Empty)*	This study
FusP-SC59	*XI-2::(Empty)*	This study
FusP-SC60	*X-3::(Empty) XI-2::(Empty)*	This study

Table S4.4 Averages of final OD_{600} of cultivations of strains expressing genetic constructs. The standard deviations are given in parenthesis.

Strain	**Avg. OD_{600} (St. Dev)**
FusP-Sc01 - WPavg	9.75 (0.47)
FusP-Sc02	9.53 (0.93)
FusP-Sc03	9.67 (0.18)
FusP-Sc04	9.63 (0.39)
FusP-Sc05	10.45 (1.55)
FusP-Sc06	9.32 (0.2)
FusP-Sc07	9.56 (0.35)
FusP-Sc08	9.41 (0.31)
FusP-Sc09	9.08 (0.89)
FusP-Sc10	9.12 (0.08)
FusP-Sc11	8.97 (0.12)
FusP-Sc12	8.83 (0.36)
FusP-Sc13	10.31 (0.12)
FusP-Sc14	10.24 (0.47)
FusP-Sc15	10.25 (0.29)
FusP-Sc16	10.52 (0.18)
FusP-Sc17	9.75 (0.41)
FusP-Sc18	8.4 (0.47)
FusP-Sc19	9.88 (0.11)
FusP-Sc20	9.73 (0.23)
FusP-Sc21	9.57 (0.22)
FusP-Sc22	9.83 (0.06)
FusP-Sc23	9.43 (0.59)
FusP-Sc24	10.27 (0.37)
FusP-Sc25	9.8 (0.07)
FusP-Sc26	9.49 (0.26)
FusP-Sc27	9.6 (0.11)
FusP-Sc28	9.52 (0.86)
FusP-Sc29	9.09 (0.38)
FusP-Sc30	9.72 (0.26)
FusP-Sc31	10.08 (0.21)
FusP-Sc32	9.32 (0.38)
FusP-Sc33	9.76 (0.52)
FusP-Sc34	9.72 (0.22)
FusP-Sc35	9.43 (0.5)
FusP-Sc36	9.63 (0.31)
FusP-Sc37	9.63 (0.37)
FusP-Sc38	9.61 (0.26)
FusP-Sc39	9.41 (0.21)
FusP-Sc40	9.75 (0.24)
FusP-Sc41	9.31 (0.18)
FusP-Sc42	10.4 (0.32)
FusP-Sc43	10.04 (0.14)
FusP-Sc44	10.19 (0.74)
FusP-Sc45	10.32 (0.35)
FusP-Sc46	7.83 (0.46)
FusP-Sc47	9.77 (0.2)
FusP-Sc48	9.59 (0.23)
FusP-Sc49	8.96 (0.07)
FusP-Sc50	6.36 (0.11)
FusP-Sc51	10.03 (0.2)
FusP-Sc52	10.36 (0.16)
FusP-Sc53	10.16 (0.08)
FusP-Sc54	10.32 (0.11)
FusP-Sc55	10.24 (0.18)
FusP-Sc56	10.29 (0.02)
FusP-Sc57	10.52 (0.08)
FusP-Sc58	10.72 (1.04)
FusP-Sc59	9.96 (0.2)
FusP-Sc60	10.16 (0.07)

Table S4.5 The whole metabolomics dataset is available for download at: https://files.dtu.dk/u/5NKLSHNSi0j20otL/Metabolomics%20dataset.xlsx?l

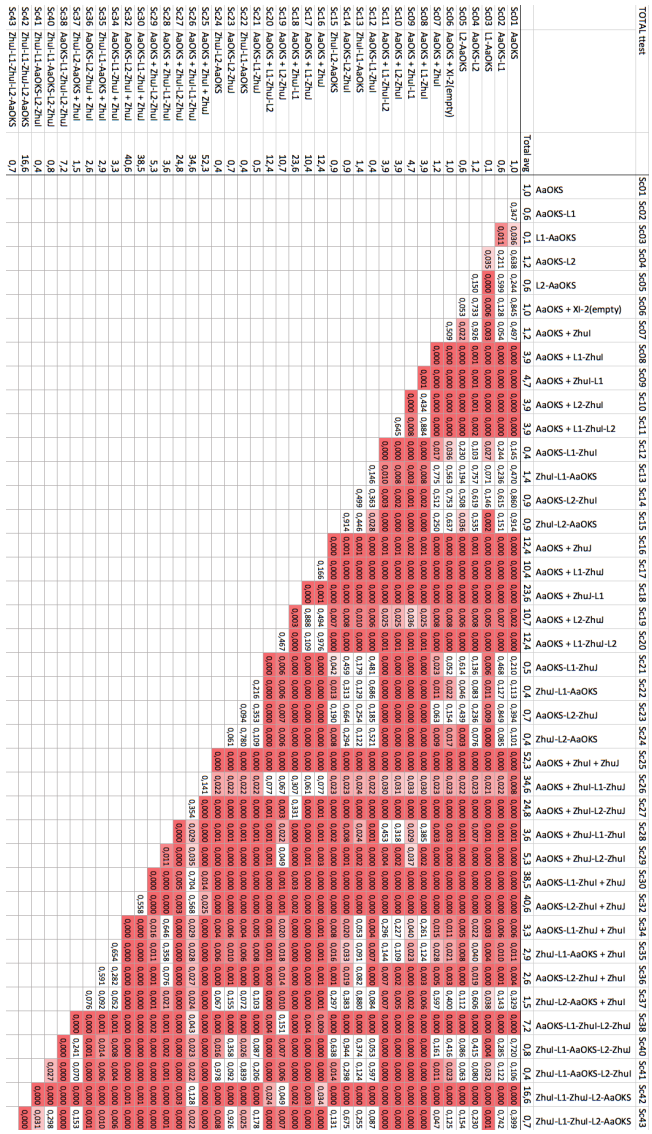


Figure S4.1 Calculated p-values when comparing the total level of quantified metabolites for each strain. The t-test was done as a two-tailed test and assumed to be of unpaired with equal variance type. Values highlighted in red are significant ($P < 0.05$).

Figure S4.2 Calculated p-values when comparing the level of flavokermesic acid for each strain. The t-test was done as a two-tailed test and assumed to be of unpaired with equal variance type. Values highlighted in red are significant ($P < 0.05$).

Figure S4.3 Calculated p-values when comparing the percentage of detected metabolite having the C7-C12 folding pattern for each strain. The t-test was done as a two-tailed test and assumed to be of unpaired with equal variance type. Values highlighted in red are significant ($P < 0.05$).

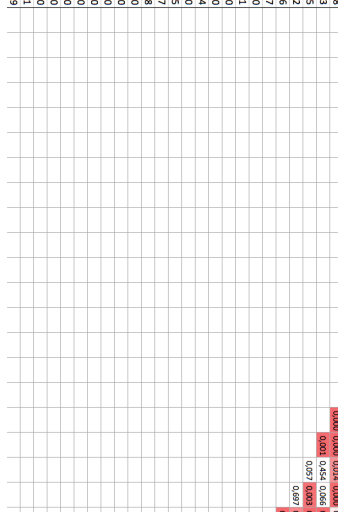


Figure S4.4 Calculated p-values when comparing the percentage of all detected metabolites being flavokermesic acid. The t-test was done as a two-tailed test and assumed to be of unpaired with equal variance type. Values highlighted in red are significant ($P < 0.05$).

Table S4.6 The whole proteomics dataset is available for download at: https://files.dtu.dk/u/opOUcjUTlHWwk5Rx/Proteomics%20dataset.xlsx?l

The sequences added to the proteome search file:

>AaOKS
MSSLSNASHLMEDVQGIRKAQRADGTATVMAIGTAHPPHIFPQDTYADFYFRATNSEHKVELKKKFDRICKKTMIGKRYFNYDEEFLKKYPNITS
FDEPSLNDRQDICVPGVPALGAEAAVKAIAEWGRPKSEITHLVFCTSCGVDMPSADFQCAKLLGLRTNVNKYCVYMQGCYAGGTVMRYAKDLAEN
NRGARVLVVCAELTIIGLRGPNESHLDNAIGNSLFGDGAAALIVGSDPIIGVEKPMFEIVCAKQTVIPNSEDVIHLHMREAGLMFYMSKDSPETI
SNNVEACLVDVFKSVGMTPPEDWNSLFWIPHPGGRAILDQVEAKLKLRPEKFRATRTVLWDCGNMVSACVLYILDEMRRKSADEGLETYGEGLEW
GVLLGFGPGMTVETILLHSLPLM

>ZhuI
MRHVEHTVTVAAPADLVWEVLADVLGYADIFPPTEKVEILEEGQGYQVVRLHVDVAGEINTWTSRRDLDPARRVIAYRQLETAPIVGHMSGEWRA
FTLDAERTQLVLTHDFVTRAAGDDGLVAGKLTPDEAREMLEAVVERNSVADLNAVLGEAERRVRAAGGVGTVTA

> ZhuJ
MSGRKTFLDLSFATRDTPSEATPVVVDLLDHVTGATVLGLSPEDFPDGMAISNETVTLTTHTGTHMDAPLHYGPLSGGVPAKSIDQVPLEWCYGP
GVRLDVRHVPAGDGITVDHLNAALDAAEHDLAPGDIVMLWTGADALWGTREYLSTFPGLTGKGTQFLVEAGVKVIGIDAWGLDRPMAAMIEEYRR
TGDKGALWPAHVYGRTREYLQLEKLNNLGALPGATGYDISCFPVAVAGTGAGWTRVVAVFEQEEED

7.3 Appendix – Manuscript III

Table S5.1 The AntiSMASH results are available for download at: https://files.dtu.dk/u/qxzA2jycwNy5HMDZ/AntiSMASH_results.xlsx?l

Figure S5.1 The constructed phylogenetic tree is available for download at: https://files.dtu.dk/u/mLoJLA86CSiKugkv/Phylogenetic_tree%28whole_dataset%29.pptx?l

Table S5.2 List of strains in the subset used in this study. NCBI accession numbers are given in parentheses.

NRRL[a]	Other[b]	Species	Geographic origin	Source	citation
2903	CBS 110244	*F. austroamericanum*	Brazil	NCAUR-MiSeq	
5538	IMI 58289	*F. fujikuroi*	Taiwan	NCBI (HF679023)	[3]
13374	FRC R-5128	*F. longipes*	New Guinea	NCAUR-MiSeq	[4]
13412	FRC Rd-34	*F. decemcellulare*	Dominican Republic	NCAUR-Ion Torrent	[4]
13444	-	*F. chlamydosporum*	Australia	NCAUR-MiSeq	[4]
20692	CBS 737.79	*F. graminum*	Ethiopia	NCAUR-MiSeq	[4]
20693	CBS 720.79	*F. heterosporum*	Netherlands	NCAUR-MiSeq	[4]
20956	CBS 123670	*F. verticillioides*	USA	Broad, MIPS	[4]
26754	CBS 110254	*F. acacia-mearnsii*	South Africa	NCAUR-MiSeq	This study
28387		*F. commune*	Netherlands	NCAUR-Ion Torrent and MiSeq	[4]
31041	Li #95	*F. virguliforme*	USA	NCAUR-MiSeq	[4]
31084	PH1 = CBS 123657	*F. graminearum*	USA	Broad, MIPS	[5]
34936	CBS 123668	*F. oxysporum* f. sp. *Lycopersici*	Spain	Broad, MIPS	[4]
36148	CBS 109638	*F. buxicola*	Belgium	NCAUR-MiSeq	[4]
45880	VanEtten 77-13-4	*F. solani* f. sp. *pisi*	USA	JGI	[6]
53430	-	*F. sibiricum*	Russia	NCAUR-MiSeq	This study
54050		*F. palustre*	USA	NCAUR-MiSeq	This study
54939	Fa5001	*F. avenaceum*	Finland	NCBI	[4]
62612	CS3096	*F. pseudograminearum*	Australia	NCBI (NC_031954)	[7]
62623	CS7071	*F. culmorum*	Australia	NCBI	This study
62941	IMI 351954	*Fusarium* sp.	Singapore	NCAUR-MiSeq	This study
-	ET1	*F. proliferatum*	Russia	NCBI (FJOF01000002)	[8]

[a] NRRL, ARS Culture Collection, Peoria, IL.

[b] CBS-KNAW, Centraalbureau voor Schimmelcultures—Fungal Biodiversity Center, Utrecht, Netherlands; FRC, Fusarium Research Center, The Pennsylvania State University, State College, PA and IMI, CABI Biosciences, Egham, Surrey, England.

	Seq length	*F. acacia-mearnsii*	*F. austroamericanum*	*F. pseudograminearum*	*F. graminearum*	*F. culmorum*	*F. sibiricum*	*F. palustre*	*F. longipes*	*F. chlamydosporum*	*F. fujikuroi*	*F. proliferatum*	*F. oxysporum*	*F. commune*	*F. verticilloides*	*F. graminum*	*F. heterosporum*	*F. avenaceum*	*Fusarium* sp. 62941	*F. solani*	*F. virguliforme*	*F. decemcellulare*	*F. buxicola*
F. acacia-mearnsii	1232		95,5	95,2	94,0	95,2	87,1	87,0	80,8	80,4	69,3	69,8	71,5	70,3	71,0	68,6	69,4	70,1	66,4	66,6	65,5	66,7	64,5
F. austroamericanum	1227	1177		95,6	92,4	93,1	87,0	87,4	81,4	80,1	68,5	69,0	70,7	69,8	70,0	68,4	68,9	69,6	65,5	66,1	64,9	66,4	64,5
F. pseudograminearum	1227	1173	1173		91,4	92,4	87,7	88,2	82,2	81,7	69,6	69,8	71,1	70,6	70,3	68,6	69,5	70,2	67,6	67,2	66,2	66,4	64,7
F. graminearum	1232	1158	1138	1126		97,2	85,2	85,4	80,0	79,6	69,0	69,3	70,7	69,3	70,4	68,5	68,7	69,5	64,9	65,5	64,3	65,8	63,7
F. culmorum	1232	1173	1147	1138	1197		85,7	85,9	80,7	80,4	69,1	69,7	71,0	69,6	70,2	68,7	69,3	70,1	65,7	66,2	65,1	66,8	63,6
F. sibiricum	1228	1073	1072	1081	1050	1056		91,7	81,8	79,6	69,4	69,7	71,4	70,8	70,9	69,7	68,9	69,3	66,3	66,6	64,7	66,7	64,4
F. palustre	1229	1072	1077	1086	1052	1058	1127		83,1	79,8	70,1	70,3	71,6	70,7	70,8	69,6	69,0	68,9	66,9	66,6	65,7	66,7	65,2
F. longipes	1218	996	1000	1010	986	994	1008	1024		77,8	69,2	69,7	70,7	69,9	70,0	67,6	67,6	69,1	66,5	67,0	65,3	67,5	65,6
F. chlamydosporum	1227	995	988	1007	985	995	985	987	959		68,8	69,1	69,3	68,8	69,0	70,1	70,3	69,8	66,4	66,5	65,9	66,9	64,6
F. fujikuroi	1212	859	849	863	856	857	858	867	856	854		97,3	91,3	90,4	88,6	70,5	72,5	71,4	66,1	67,1	67,2	68,1	65,8
F. proliferatum	1212	866	855	865	859	864	862	870	862	858	1179		91,3	90,5	89,3	70,4	72,5	71,8	66,1	66,9	67,3	67,8	65,7
F. oxysporum	1213	886	876	882	877	880	883	886	874	860	1108	1107		93,7	90,4	71,6	72,8	71,3	66,7	67,4	67,1	68,8	66,3
F. commune	1213	872	866	875	859	863	876	874	864	854	1097	1098	1137		88,6	71,2	73,0	71,7	66,6	67,1	67,2	67,8	66,5
F. verticilloides	1220	880	868	872	873	871	879	878	866	856	1081	1089	1103	1081		71,7	72,6	72,2	66,8	67,2	68,0	67,8	66,3
F. graminum	1206	848	845	848	847	849	862	860	834	867	860	859	873	868	875		89,0	79,1	66,2	66,1	66,6	66,6	66,1
F. heterosporum	1209	858	851	859	849	856	851	853	834	869	884	884	888	891	886	1076		80,5	66,5	66,9	67,2	68,6	67,5
F. avenaceum	1215	866	860	867	859	866	857	852	852	863	872	877	871	875	881	961	978		68,1	68,0	68,2	68,5	66,7
Fusarium sp. 62941	1227	833	821	847	815	825	832	839	833	832	827	827	834	833	835	827	830	853		89,1	82,6	74,9	73,4
F. solani	1236	837	829	842	823	831	836	836	840	835	840	838	844	840	841	828	838	851	1101		82,3	76,5	74,5
F. virguliforme	1239	825	816	833	810	820	815	828	822	830	844	845	843	844	854	836	844	857	1026	1022		73,4	73,8
F. decemcellulare	1238	836	830	831	824	837	834	834	844	838	849	845	858	845	847	832	857	856	932	951	916		75,5
F. buxicola	1256	817	815	818	807	805	815	826	829	817	832	831	838	840	838	835	853	843	924	938	931	950	

Identity%: 63,6% 97,3%
Identity count: 805 1197

Figure S5.2 Plot showing the identity of the different *pgl2* genes from different species within the *Fusarium* genus.

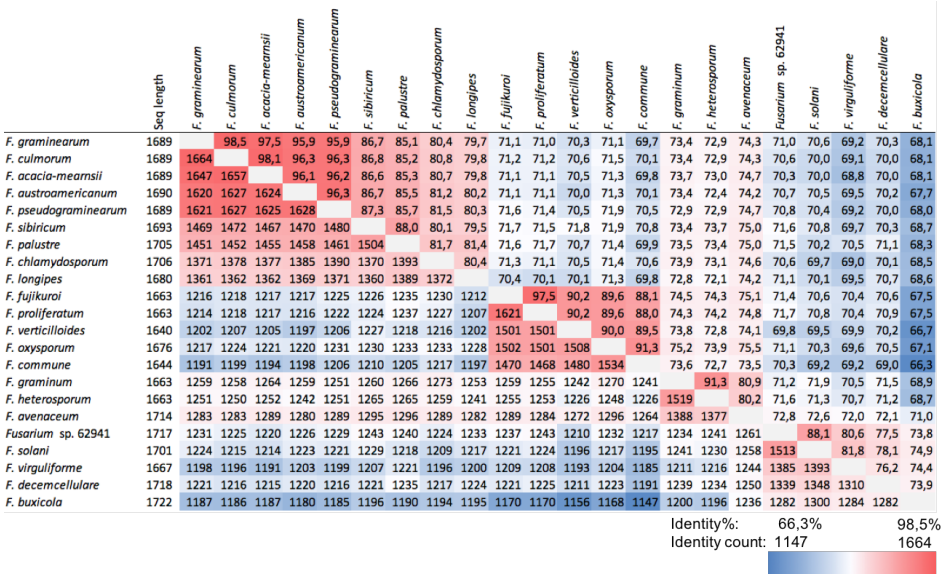


Figure S5.3 Plot showing the identity of the different *pgl3* genes from different species within the *Fusarium* genus.

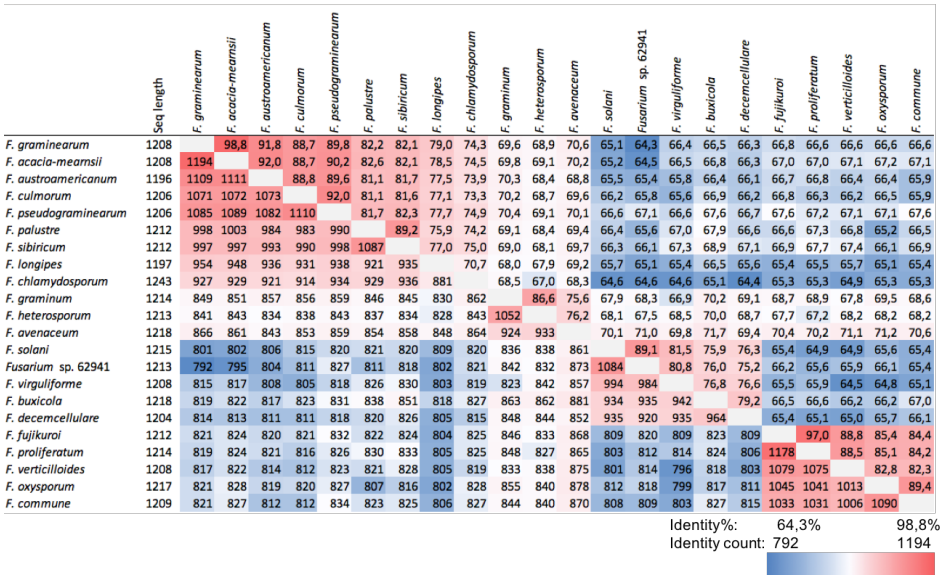


Figure S5.4 Plot showing the identity of the different *pgl4* genes from different species within the *Fusarium* genus.

	Seq length	*F. graminearum*	*F. acacia-mearnsii*	*F. austroamericanum*	*F. culmorum*	*F. pseudograminearum*	*F. palustre*	*F. sibiricum*	*F. longipes*	*F. chlamydosporum*	*F. graminum*	*F. heterosporum*	*F. avenaceum*	*F. fujikuroi*	*F. proliferatum*	*F. oxysporum*	*F. commune*	*F. verticilloides*	*F. solani*	*Fusarium* sp. 62941	*F. virguliforme*	*F. decemcellulare*	*F. buxicola*
F. graminearum	801		99,1	98,8	94,4	91,6	86,6	87,5	84,3	84,1	79,4	78,5	81,5	77,7	77,5	77,2	77,9	77,8	72,5	72,7	74,2	75,2	75,1
F. acacia-mearnsii	801	794		99,6	95,3	92,5	87,5	88,2	84,8	84,5	80,0	79,2	81,9	77,9	77,8	77,4	78,2	78,3	72,4	72,9	74,1	75,3	75,1
F. austroamericanum	801	791	798		94,9	92,1	87,1	87,8	84,6	84,4	79,9	79,0	81,5	78,0	77,7	77,3	78,0	78,4	72,4	72,7	74,0	74,9	74,7
F. culmorum	801	756	763	760		92,0	87,3	87,9	85,6	83,4	80,9	79,9	81,3	77,9	77,7	77,3	77,8	78,8	72,5	72,2	74,1	75,4	75,3
F. pseudograminearum	795	734	741	738	737		86,0	86,5	83,4	83,0	79,0	78,5	79,7	77,4	76,9	77,5	77,9	78,0	71,4	70,8	73,9	74,9	74,2
F. palustre	801	694	701	698	699	689		90,5	85,5	84,8	77,7	77,4	79,0	76,2	76,2	76,7	76,5	76,8	72,6	72,0	72,1	75,8	75,7
F. sibiricum	801	702	707	704	705	694	726		85,8	84,4	78,3	77,7	79,4	76,1	75,9	76,8	76,4	76,6	72,0	72,0	72,3	75,2	75,2
F. longipes	801	675	679	678	686	668	685	688		86,1	81,4	81,2	80,5	79,0	78,5	78,8	79,4	79,5	74,4	74,1	74,0	73,7	75,7
F. chlamydosporum	801	674	677	676	668	665	679	677	690		78,9	78,3	80,0	77,3	77,0	78,2	77,9	77,0	73,1	73,7	72,5	74,6	74,7
F. graminum	801	636	641	640	648	633	622	628	652	632		94,8	87,6	80,0	79,8	78,9	80,4	80,8	74,4	72,6	74,6	75,4	75,4
F. heterosporum	801	629	634	633	640	629	620	623	650	627	759		86,6	79,7	79,2	78,7	79,7	80,3	73,0	71,8	73,7	75,4	75,6
F. avenaceum	801	653	656	653	651	638	633	637	645	641	702	694		81,5	81,0	80,7	81,4	81,7	74,5	73,7	74,4	76,7	77,9
F. fujikuroi	798	622	624	625	624	620	610	610	633	619	641	638	653		97,1	91,7	92,9	93,0	74,0	74,0	73,9	76,2	76,9
F. proliferatum	798	621	623	622	622	616	610	609	629	617	639	634	649	775		91,6	92,7	92,5	74,6	73,9	73,7	76,7	76,6
F. oxysporum	798	618	620	619	619	621	614	616	631	626	632	630	646	732	731		94,1	90,7	75,2	74,7	74,4	76,5	77,7
F. commune	798	624	626	625	623	624	613	613	636	624	644	638	652	741	740	751		92,1	74,6	74,4	74,1	76,3	76,9
F. verticilloides	798	623	627	628	631	625	615	614	637	617	647	643	654	742	738	724	735		75,7	75,5	74,6	76,5	78,1
F. solani	807	585	584	584	585	576	586	582	600	590	600	589	601	597	602	607	602	611		93,4	88,6	78,8	77,0
Fusarium sp. 62941	807	587	588	587	583	571	581	582	598	595	586	579	595	597	596	603	600	609	754		86,6	78,4	77,7
F. virguliforme	807	599	598	597	598	596	582	584	597	585	602	595	600	596	595	600	598	602	715	699		80,1	78,3
F. decemcellulare	798	603	604	601	605	601	608	604	591	598	605	605	615	610	614	613	611	613	636	633	646		81,9
F. buxicola	801	602	602	599	604	595	607	604	607	599	605	606	625	617	614	623	617	626	621	627	632	656	

Identity%: 70,8% 99,6%
Identity count: 571 798

Figure S5.5 Plot showing the identity of the different *pgl5* genes from different species within the *Fusarium* genus.

	Seq length	*F. acacia-mearnsii*	*F. austroamericanum*	*F. graminearum*	*F. pseudograminearum*	*F. culmorum*	*F. palustre*	*F. sibiricum*	*F. longipes*	*F. chlamydosporum*	*F. fujikuroi*	*F. proliferatum*	*F. commune*	*F. oxysporum*	*F. verticilloides*	*F. avenaceum*	*F. graminum*	*F. heterosporum*	*F. solani*	*Fusarium* sp. 62941	*F. virguliforme*	*F. buxicola*	*F. decemcellulare*
F. acacia-mearnsii	1119		98,9	98,6	93,9	94,0	82,3	83,0	69,6	70,7	56,4	56,5	56,7	56,2	55,4	58,7	53,0	52,7	51,6	52,6	52,8	54,2	52,1
F. austroamericanum	1119	1107		98,6	93,9	94,0	82,2	83,5	69,6	71,1	56,1	56,1	56,6	56,2	55,2	58,6	52,9	52,6	51,5	52,6	52,5	53,8	51,7
F. graminearum	1119	1103	1103		93,9	94,0	82,3	83,2	69,5	71,0	56,2	56,5	56,7	56,2	55,4	59,0	53,1	52,9	51,9	52,5	52,8	54,6	52,1
F. pseudograminearum	1119	1051	1051	1051		92,3	82,0	82,2	68,2	70,2	55,8	56,2	56,2	55,6	54,6	59,1	53,0	52,4	51,4	51,8	52,3	54,0	51,6
F. culmorum	1119	1052	1052	1052	1033		81,9	82,3	69,8	70,2	56,9	56,8	57,4	56,5	55,5	58,8	52,5	51,9	51,6	51,5	53,0	54,1	52,6
F. palustre	1119	921	920	921	918	916		85,7	68,2	68,8	55,5	55,2	55,6	55,0	54,8	57,3	52,6	52,6	50,5	50,4	51,2	53,3	51,4
F. sibiricum	1119	929	934	931	920	921	959		68,9	68,9	55,3	55,3	55,6	55,6	54,5	57,7	53,0	52,4	51,2	51,7	51,7	54,4	52,1
F. longipes	1101	791	791	790	775	794	775	783		66,8	52,3	52,0	52,3	53,0	52,0	56,1	49,5	49,7	50,0	50,7	49,7	52,2	49,1
F. chlamydosporum	1149	812	817	816	806	807	790	792	774		56,2	56,4	56,2	55,9	55,4	60,8	54,1	52,9	51,9	52,7	51,5	55,6	52,1
F. fujikuroi	1017	661	657	659	654	667	650	648	616	664		96,9	92,1	90,5	89,1	65,9	56,3	55,8	50,1	49,5	52,2	53,8	52,5
F. proliferatum	1020	662	658	662	659	666	647	648	612	666	988		91,6	90,3	88,6	66,3	56,5	55,8	49,6	49,2	51,3	53,9	52,2
F. commune	1008	665	663	665	659	673	651	652	616	664	937	934		92,5	88,8	66,2	55,4	55,7	49,7	49,1	51,6	53,7	52,2
F. oxysporum	1008	659	659	659	651	662	644	651	624	660	920	921	932		88,0	65,6	55,2	55,7	49,0	48,6	50,9	52,6	51,5
F. verticilloides	990	647	645	647	638	649	640	637	611	652	906	904	898	890		64,9	55,8	55,7	48,3	48,1	50,9	52,3	51,2
F. avenaceum	1152	705	704	708	710	706	688	693	681	734	760	764	763	756	748		65,5	65,5	54,9	53,6	53,2	58,7	57,5
F. graminum	1128	636	635	638	636	630	632	636	599	653	650	652	640	638	644	762		86,0	50,0	48,2	48,8	52,2	52,2
F. heterosporum	1128	633	632	635	629	623	632	629	602	639	645	645	643	643	643	762	970		49,9	49,0	47,9	52,4	51,8
F. solani	1167	631	629	634	628	630	617	626	616	637	605	599	600	592	584	670	610	608		91,1	80,1	70,3	69,6
Fusarium sp. 62941	1158	638	638	637	628	625	611	627	619	642	598	595	594	588	582	654	587	597	1066		79,3	68,9	68,2
F. virguliforme	1101	632	629	632	626	635	613	619	597	621	620	609	613	604	604	651	595	584	940	923		67,0	67,4
F. buxicola	1188	659	655	665	657	658	649	662	640	679	651	652	650	636	633	715	635	637	846	824	801		73,5
F. decemcellulare	1179	638	633	638	632	644	630	638	606	641	632	629	629	620	617	698	631	627	834	817	805	882	

Identity%: 47,9% 98,9%
Identity count: 582 1107

Figure S5.6 Plot showing the identity of the different *pgl6* genes from different species within the *Fusarium* genus.

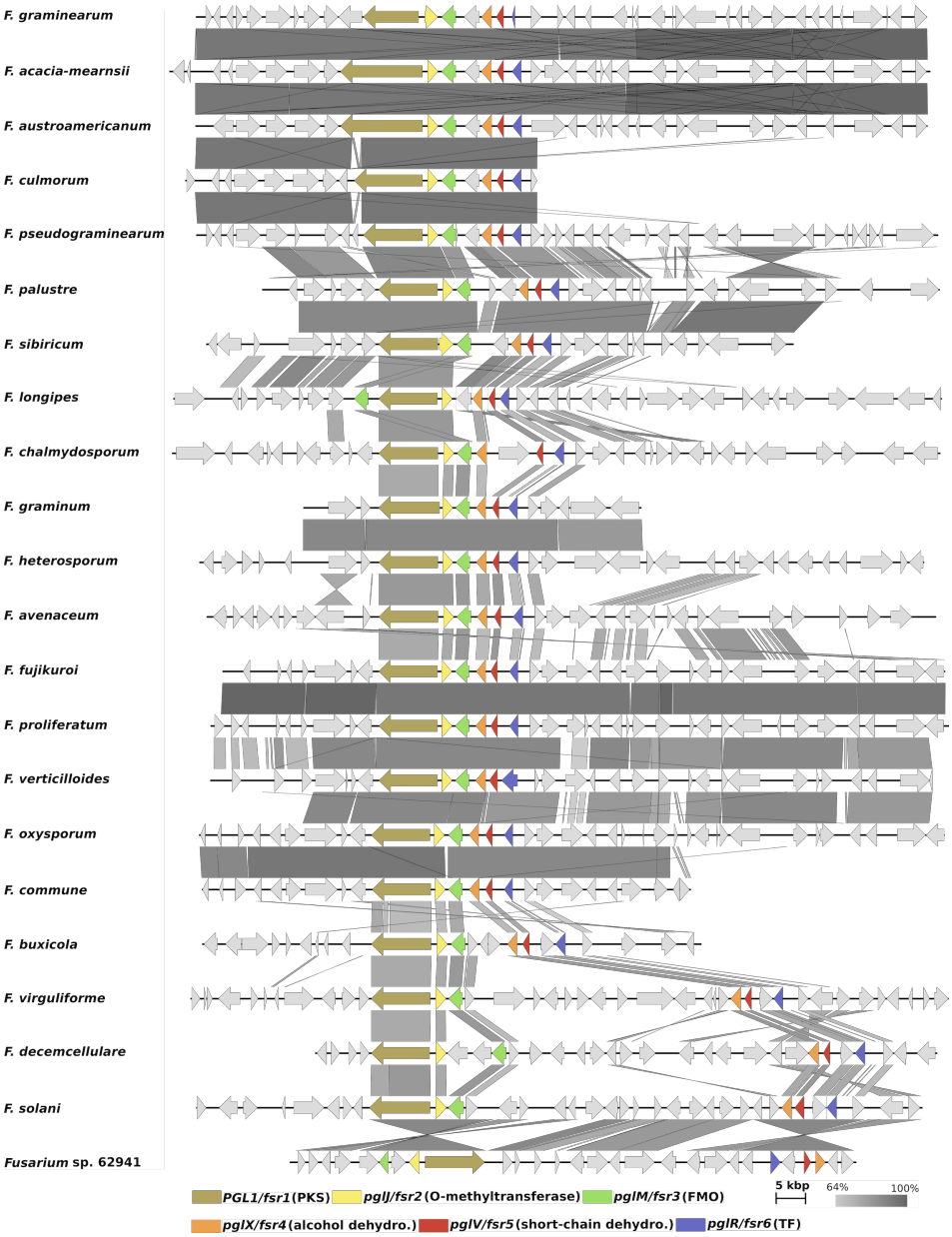


Figure S5.7 Synteny of the genetic locus for the *PGL1* cluster.

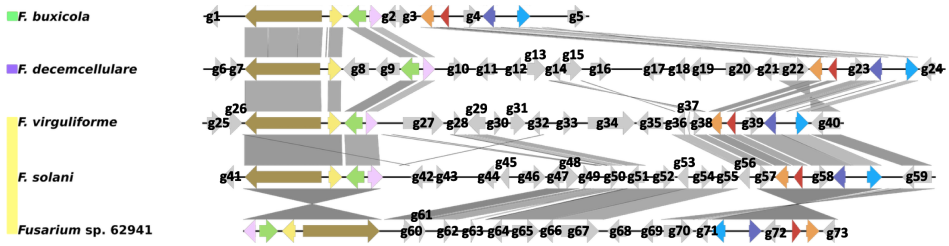


Figure S5.8 Numbering of genes in the extra genes of species complexes *buxicola*, *decemcellulare* and *solani* of investigated isolates.

Table S5.3 List of putative function for extra genes in species complexes *buxicola*, *decemcellulare* and *solani* of investigated isolates. Numbering of genes are as in figure S5.8.

Gene	Isolate	Length (AA)	PFAM Hit	Domains	Predicted function
1	F. buxicola	278	PF05572	Peptidase_M43	
2	F. buxicola	250	-	-	-
3	F. buxicola	398	PF00891	Methyltransferase	Methyltransferase
4	F. buxicola	447	PF07690	MFS transporter	
5	F. buxicola	556	PF00324	AA_permease	
6	F. decelcellulare	301	PF04479	RTA1	Membrane protein
7	F. decelcellulare	347	-	-	-
8	F. decelcellulare	786	PF00004	AAA	ATPase
9	F. decelcellulare	728	PF06985	HET	Heterokaryon incompability protein
10	F. decelcellulare	489	PF02458	Transferase	Transferase
11	F. decelcellulare	510	PF00067	p450	Cytochrome P450
12	F. decelcellulare	334	PF10014	2OG-Fe_Oxy_2	Fe-dependant dioxygenase
13	F. decelcellulare	576	PF01053	Cys_Met_Meta_PP	PLP-dependant Cys/Met metabolism
14	F. decelcellulare	352	PF00106	Short chain dehydrogenase	Short chain dehydrogenase
15	F. decelcellulare	460	-	-	-
16	F. decelcellulare	566	PF01494	FAD_binding_3	FMO
17	F. decelcellulare	544	PF00083	Sugar transporter	Sugar transport
18	F. decelcellulare	343	PF00069	Pkinase	Protein kinase
19	F. decelcellulare	619	-	-	-
20	F. decelcellulare	857	PF08407 + PF01644	Chitin_synth_1	Chitin synthase
21	F. decelcellulare	461	PF00232	Glyco_hydro_1	Glycoside hydrolase
22	F. decelcellulare	896	PF05592 + PF08531	Bacterial rhamnosidase	Glycoside hydrolase
23	F. decelcellulare	427	PF07690	MFS transporter	MFS transporter
24	F. decelcellulare	482	PF07632	DUF1593	Unknown
25	F. virguliforme	503	PF13520	AA_permease_2	Amino acid permease
26	F. virguliforme	361	-	-	-
27	F. virguliforme	1269	PF00445 + PF01565	Ribonuclease_T2 + FAD_binding_4	FAD-binding RNase
28	F. virguliforme	238	-	-	-
29	F. virguliforme	550	PF02129 + PF08530	X-Pro dipeptidyl-peptidase	Protease
30	F. virguliforme	289	PF06985	HET	Heterokaryon incompability protein
31	F. virguliforme	458	-	-	-
32	F. virguliforme	564	PF06985	HET	Heterokaryon incompability protein
33	F. virguliforme	398	-	-	-
34	F. virguliforme	1131	-	-	-

35	F. virguliforme	832	PF13625 + PF04851 + PF16203	Helicase_C_3 + ResIII + ERCC3_RAD25_C	Helicase
36	F. virguliforme	291	-	-	-
37	F. virguliforme	104	-	-	-
38	F. virguliforme	315	PF00106	Short chain dehydrogenase	Short chain dehydrogenase
39	F. virguliforme	409	PF07690	MFS transporter	MFS transporter
40	F. virguliforme	889	PF05592 + PF08531	Bacterial rhamnosidase	Glycoside hydrolase
41	F. solani	338	PF00962	Adenosine/AMP Deaminase	Adenosine deaminase
42	F. solani	703	PF04082	Fungal TF	Transcription factor
43	F. solani	596	PF13561	Enoyl-(Acyl carrier protein) reductase	Enoyl reductase
44	F. solani	286	PF08240 + PF00107	Alcohol dehydrogenase + Zinc binding dehydrogenase	Dehydrogenase
45	F. solani	731	PF13561	Enoyl-(Acyl carrier protein) reductase	Enoyl reductase
46	F. solani	538	PF00083	Sugar transporter	Sugar transport
47	F. solani	342	PF13489	Methyltransferase	Methyltransferase
48	F. solani	378	PF01370	Epimerase	NADH dehydrogenase (Ubiquinone)
49	F. solani	165	-	-	-
50	F. solani	248	-	-	-
51	F. solani	550	PF02129 + PF08530	X-Pro dipeptidyl-peptidase	Protease
52	F. solani	667	PF04082	Fungal TF	Transcription factor
53	F. solani	332	PF05544	Proline racemase	Proline racemase
54	F. solani	477	PF07690	MFS transporter	MFS transporter
55	F. solani	350	-	-	-
56	F. solani	396	PF11951	Fungal_trans_2	Fungal TF
57	F. solani	344	PF00106	Short chain dehydrogenase	Short chain dehydrogenase
58	F. solani	439	PF07690	MFS transporter	MFS transporter
59	F. solani	896	PF05592 + PF08531	Bacterial rhamnosidase	Glycoside hydrolase
60	Fusarium sp. 62941	377	PF01370	Epimerase	NADH dehydrogenase (Ubiquinone)
61	Fusarium sp. 62941	170	-	-	-
62	Fusarium sp. 62941	247	-	-	-
63	Fusarium sp. 62941	179	-	-	-
64	Fusarium sp. 62941	551	PF02129 + PF08530	X-Pro dipeptidyl-peptidase	Protease
65	Fusarium sp. 62941	660	PF04082	Fungal TF	Transcription factor
66	Fusarium sp. 62941	403	PF05544	Proline racemase	Proline racemase
67	Fusarium sp. 62941	850	PF07690	MFS transporter	MFS transporter
68	Fusarium sp. 62941	539	PF00083	Sugar transporter	Sugar transport
69	Fusarium sp. 62941	506	PF00232	Glycoside hydrolase	Glycoside hydrolase
70	Fusarium sp. 62941	896	PF05592 + PF08531	Bacterial rhamnosidase	Glycoside hydrolase
71	Fusarium sp. 62941	358	-	-	-
72	Fusarium sp. 62941	439	PF07690	MFS transporter	MFS transporter
73	Fusarium sp. 62941	345	PF00106	Short chain dehydrogenase	Short chain dehydrogenase

Additional references

[1] Mikkelsen, MD, Buron, LD, Salomonsen, B, Olsen, CE, Hansen, BG, Mortensen, UH, and Halkier, BA. "Microbial production of indolylglucosinolate through engineering of a multi-gene pathway in a versatile yeast expression platform". In: *Metab Eng* 14.2 (2012), pp. 104–111.

[2] Entian, KD and Kötter, P. "25 Yeast Genetic Strain and Plasmid Collections". In: *Methods Microbiol* 36.06 (2007), pp. 629–666.

[3] Wiemann, P, Sieber, CM, Von Bargen, KW, Studt, L, Niehaus, EM, Espino, JJ, Huß, K, Michielse, CB, Albermann, S, Wagner, D, et al. "Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites". In: *PLoS Pathogens* 9.6 (2013), pp. 1–35.

[4] O'Donnell, K, Rooney, AP, Proctor, RH, Brown, DW, McCormick, SP, Ward, TJ, Frandsen, RJN, Lysøe, E, Rehner, SA, Aoki, T, Robert, VA, Crous, PW, Groenewald, JZ, Kang, S, and Geiser, DM. "Phylogenetic analyses of RPB1 and RPB2 support a middle Cretaceous origin for a clade comprising all agriculturally and medically important fusaria". In: *Fungal Genet Biol* 52 (2013), pp. 20–31.

[5] Cuomo, CA, Güldener, U, Xu, JR, Trail, F, Turgeon, BG, Di Pietro, A, Walton, JD, Ma, LJ, Baker, SE, Rep, M, et al. "The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization". In: *Science* 317.5843 (2007), pp. 1400–1402.

[6] Coleman, JJ, Rounsley, SD, Rodriguez-Carres, M, Kuo, A, Wasmann, CC, Grimwood, J, Schmutz, J, Taga, M, White, GJ, Zhou, S, et al. "The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion". In: *PLoS genetics* 5.8 (2009), e1000618.

[7] Gardiner, DM, McDonald, MC, Covarelli, L, Solomon, PS, Rusu, AG, Marshall, M, Kazan, K, Chakraborty, S, McDonald, BA, and Manners, JM. "Comparative pathogenomics reveals horizontally acquired novel virulence genes in fungi infecting cereal hosts". In: *PLoS Pathogens* 8.9 (2012), pp. 1–22.

[8] Tsavkelova, EA, Bömke, C, Netrusov, AI, Weiner, J, and Tudzynski, B. "Production of gibberellic acids by an orchid-associated *Fusarium proliferatum* strain". In: *Fungal Genet Biol* 45.10 (2008), pp. 1393–1403.